Aalto University
School of Science
and Technology

Sanna Syrjänen

The effects of visual processing on human frequency following brain response

In partial fulfillment of the requirements for the degree of Master of Science, Espoo May 6th, 2010.

Supervisor: Prof. Mikko Sams

Instructor: M.Sc. Jaakko Kauramäki

AALTO YLIOPISTO **DIPLOMITYÖN**
TEKNILLINEN KORKEAKOULU **TIIVISTELMÄ**

Tekijä: Sanna Syrjänen

Otsikko: Visuaalisen prosessoinnin vaikutus ihmisaivojen taajuusseuraajavasteeseen

Päivämäärä: 6. toukokuuta, 2010 **Sivumäärä:** 54

Osasto: Lääketieteellisen tekniikan ja laskennallisen tieteen laitos
Professuuri: S-114, Laskennallinen tekniikka

Työn valvoja: Prof. Mikko Sams
Työn ohjaaja: DI Jaakko Kauramäki

Tiivistelmäteksti:

Tässä työssä tutkittiin, miten huomion keskittäminen visuaaliseen ärsykkeeseen vaikuttaa aikaiseen kuuloprosessointiin elektroenkefalografian (EEG) avulla. Erityisesti haluttiin selvittää vaikutukset kuuloaivorunkovasteen taajuusseuraajavasteeseen.

EEG mitattiin koehenkilöiden keskittyessä seuraamaan visuaalisia ärsykkeitä samalla, kun taustalla toistettiin /da/-ääniärsykettä. Kokeessa oli kolme eri tilannetta, jotka erosivat toisistaan visuaalisen ärsykkeen osalta. (1) Vokaalit, jossa koehenkilöt katsoivat äänetöntä videota vokaaleja toistavasta henkilöstä. (2) Laajenevat renkaat, jossa koehenkilöiden katsomalla videolla henkilön suun alueella liikkui laajeneva ja supistuva rengas, jolla luotiin ajallisesti ja avaruudellisesti vokaali-tilannetta vastannut liikkeen havainnointi ilman kielellistä sisältöä. (3) Staattinen kuva, jossa koehenkilölle näytettiin pysähtynyttä neutraalia kasvokuvaa. Vokaalit- ja laajenevat renkaat - tilanteissa koehenkilöiden tehtävänä oli reagoida kahteen peräkkäiseen samaan vokaaliin / renkaan suuntaan. Koehenkilöt suorittivat kunkin tehtävän kaksi kertaa satunnaisessa järjestyksessä, jotta aivorunkovasteen toistettavuutta voitiin seurata.

Koehenkilöt tunnistivat vokaalit-tilanteen kohdeärsykkeet prosentuaalisesti paremmin, mutta hitaammin, kuin laajenevat renkaat -tilanteessa. EEG tuloksissa on nähtävissä trendi, jossa taajuusseuraajavasteen amplitudi on vokaalit-tilanteessa alhaisempi. Taajuusseuraajavasteen ensimmäisen piikin osalta tulokset olivat tilastollisesti merkitseviä. Nämä tulokset voivat viitata siihen, että huuliltaluku vaikuttaa alentavasti taajuusseuraajavasteen amplitudeihin.

Avainsanat: elektroenkefalografia, audiovisuaalinen integraatio, auditorinen aivorunkovaste, taajuusseuraajavaste

AALTO UNIVERSITY **ABSTRACT OF THE**
SCHOOL OF SCIENCE AND TECHNOLOGY **MASTER'S THESIS**

Author:	Sanna Syrjänen
Title:	The effects of visual processing on human frequency following brain response
Date:	May 6th, 2010 **Number of pages:** 54
Department:	Department of Biomedical Engineering and Computational Science
Professorship:	S-114, Computational technology
Supervisor:	Prof. Mikko Sams
Instructor:	M.Sc.Jaakko Kauramäki

Abstract:

In this thesis, the effects of attending to visual stimuli to early auditory processing was studied with the aid of electroenchephalography (EEG). The focus was on the frequency following response (FFR) of the auditory brainstem response (ABR).

EEG-responses were measured while subjects attended to visual stimuli combined with auditory /da/-syllable. There were three different conditions based on the visual stimulus type: (1) vowels, where subjects attended to a video of a female speaker mouthing Finnish vowels; (2) expanding rings, where the video consisted of the same female speaker's still face with an expanding ring/oval imposed over the mouth region, creating temporally and spatially similar movement to the vowels condition, but without the linguistic content; and (3) the still condition, where the still image of the same female speaker was presented. Tasks in the vowels and expanding rings conditions were to react to two consecutive same vowels/rings. Subjects completed two sets of the three conditions in randomized order to control the intra-subject ABR replicability.

The behavioral results showed the subjects identified the vowel targets better, but slower in comparison to the expanding rings. The EEG results showed a trend towards lower FFR amplitudes during lipreading/vowels condition and there were statistically significant effects in the first peak of the FFR. These results could suggest suppressive effects of lipreading on the FFR.

Keywords: electroenchephalography, auditory brainstem response, frequency following response, audiovisual integration

Foreword

This work was carried out in the Department of Biomedical Engineering and Computational Science in the Aalto University. The supervisor of this work was Professor Mikko Sams and the instructor was M.Sc. Jaakko Kauramäki.

I would like to thank my supervisor Prof. Mikko Sams for the opportunity to do this thesis and my instructor Jaakko Kauramäki for all his guidance during this process. I would also like to thank the people who were working at the Magnet house while this work was in progress, you made it fun.

Finally I would like to thank Åge-Johan, who was there for me.

In Espoo, 6.5.2010

Sanna Syrjänen

Contents

List of Figures

List of Tables

Abbreviations and notations

A	Auditory
ABR	Auditory brainstem response
AP	Action potential
AV	Audio-visual, including both auditory and visual components
CNS	Central nervous system
EEG	Electroencephalography
EoG	Electro-oculogram
ERP	Event-related potential
FFR	Frequency following response
IC	Inferior colliculus
ISI	Interstimulus interval
LGN	Lateral geniculate nucleus
MGN	Medial geniculate nucleus
N1, N100	Negative peak occurring approximately at 100 ms after stimulus onset
PNS	Peripheral nervous system
RMA	Rectified mean amplitude
S.D.	Standard deviation (statistical)
SC	Superior colliculus
S.E.M	Standard error of means (statistical)
SPL	Sound pressure level
V	Visual
V1	The primary visual cortex (Striate cortex)

Chapter 1

Introduction

Imagine watching a movie or a television series, where the original dialogue has been replaced by a new recording done in another language. The feeling that something is awry comes from the mixed messages your different senses are relaying to your brain: the actors are visibly speaking English, but what you are hearing is German. This is an example on how the perception of speech does not rely only on the audible content, but – if such are available – the visual cues as well.

Probably the most known experiment in the field of speech perception was done by McGurk and MacDonald in 1976 when they demonstrated that conflicting visual cues presented simultaneously with auditory speech stimulus causes errors in the perception of short syllables. The effect is known as the "McGurk-effect" and has been widely used in psychophysical experiments ever since (McGurk and MacDonald, 1976).

Paul Broca (in 1861) and Carl Wernicke (in 1876) made major discoveries in the localization of speech production and comprehension in the brain based on the observations they made of patients who had suffered strokes. Post mortem observations on the lesions caused by the strokes combined with observations of the patients' impaired lingual functions while they were still alive lead to the definition of Broca's and Wernicke's areas. Localization maps of the cortex have since become more and more spesific. Connections between different areas create the basis for multimodal observations of the surrounding

world – such as audiovisual speech.

The goal of this thesis is to determine whether perception of visual speech will modulate the auditory brainstem responses to a sound in the human brain. Many studies focusing on audiovisual integration have focused on the auditory cortical effects, such as Besle et al. (2004), who demonstrated suppressive effects of visual speech to N1 auditory cortex responses to speech syllables (EEG-study) and Kauramäki et al. (2010) who found suppressive effects of visual speech and covertly produced speech to the N1 magnetic counterpart (MEG-study), latter suggesting that the suppressive effects raise from top-down influences by efferent copy signal from the speech production system. Musacchia et al. (2006) found significant effects of visual speech to the auditory brainstem response to speech syllables in latency and size: visual speech caused delays and diminished size in the onset response part of the ABR to speech. Whether the lipreading-evoked suppression starts already at the lower parts of the auditory system remains to be proved. Here we studied specifically the effects of visual speech to the frequency following response portion of the brainstem response by recording the EEG activity while presenting short speech sounds and varying visual stimuli.

Chapter 2

Background

2.1 Anatomical basis

In this chapter anatomical basis of the human nervous system is shortly introduced in general based on Bear et al. (2001); Gazzaniga et al. (2002) and Mildner (2008).

2.1.1 The human nervous system

The human nervous system is divided into two sections: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain (cerebrum, cerebellum and the brainstem) and of the spinal cord. The PNS consists of all the other nerves that are located outside the scope of the CNS.

Structure of a neuron

The human nervous system consists of two classes of cells: neurons and glial cells. Neurons form connective networks that transmit electrical signals from one to another over synaptic gaps usually with the aid of neurotransmitters. Single neuron is constructed of

the cell body, soma, the axon which extends from the soma and tree-like dendrites. The axon terminates in a disk-like axon terminal, where it connects to other neuros creating synapses (the neuron is said to innervate the receiving cell). The axon bringing the signal to the synapse is referred to as presynaptic and the neuron receiving the signal as postsynaptic. Nerves consist of bundles of individual nerve fibers, and even though the flow of information in an axon is strictly one directional, the bundles consist of axons delivering information up- and downstream between the peripheral and central nervous system.

Cell membrane surrounds the whole of a single neuron, giving it its structure and separating neuronal cytoplasm inside the cell from the surrounding environment. Normally the membrane is relatively impermeable, but distributed across it are proteine-constructed ion channels, gates and pumps that regulate the flow of electrically charged ions through the neuronal membrane. In resting state neuronal cytoplasm has higher concentration of positively charged potassium ions (K^+) while the outside fluid is rich in positively charged sodium ions (Na^+), creating ionic concentration gradients across the membrane. This causes the cell to have a negative resting potential, usually about -65 mV, which is maintained by sodium-potassium pumps. For more in-depth description of the ionic basis of the resting membrane potential, see for example Bear et al. (2001, chapter 3) or Gazzaniga et al. (2002, chapter 2).

Action potential

The action potential, the signal to carry information through the nervous system, is a rapid reversal of the polarity of the neuron from the negative resting state to positive and back.

When a neuron is being stimulated, a rapid depolarization of the membrane occurs. Not all stimulation will cause an action potential – the depolarization must reach certain threshold level in voltage to trigger an actual AP. The following repolarization of the cell causes the potential to briefly drop lower than the original resting potential, making it more difficult to reach the threshold to trigger a new AP. This state is called the refractory period, and combined with the absolute refractory period, during which the voltage-gated Na^+ channels are unable to reopen cause limitations to how often a neuron can generate

an AP. The fastest neurons can reach up to 1000 Hz in firing frequency (Niedermeyer and da Silva, 1999).

The amplitude of the action potential is independent of the inducing current, as long as it reaches the threshold level. Instead the nervous system uses firing frequency to code the intensity of the stimuli. In addition, connections between neurons can be either excitatory or inhibitory and the final effect is determined by the summation of incoming signals to the cell.

When multiple neurons fire simultaniously, for instance in the auditory nerve during a auditory stimulation, the summated electrical activity is strong enough to be measured as far-field potentials on the scalp with the aid of EEG equipment (see Section 2.2.1).

An actual perception of a stimuli is at the end of a long chain of events. What is needed is a strong enough stimuli to engage the receptor neurons in the sensory organ, which translate the stimuli into electrochemical energy. Primary sensory neurons connected to the reseptor neurons in turn take care of the neural coding of the stimulus, with information about intensity, duration and location and transfer this information into the thalamus. From there, meaningful information is chosen to be relayed onwards to the primary sensory areas on the cortex, where more segregating and grouping is performed. This prefiltered information is then relayed into the secondary sensory areas, where their meaning is found and information again transferred onwards to the tertiary areas on the cortex, ready to be interpreted and associated (Mildner, 2008).

Structure of the brain

The cerebellum and the brainstem

The cerebellum is primarily concerned with movement control, having connections arriving from the brainstem for visual, auditory and vestibular information and outputs to the thalamus and from there upwards to the premotor and motor cortex.

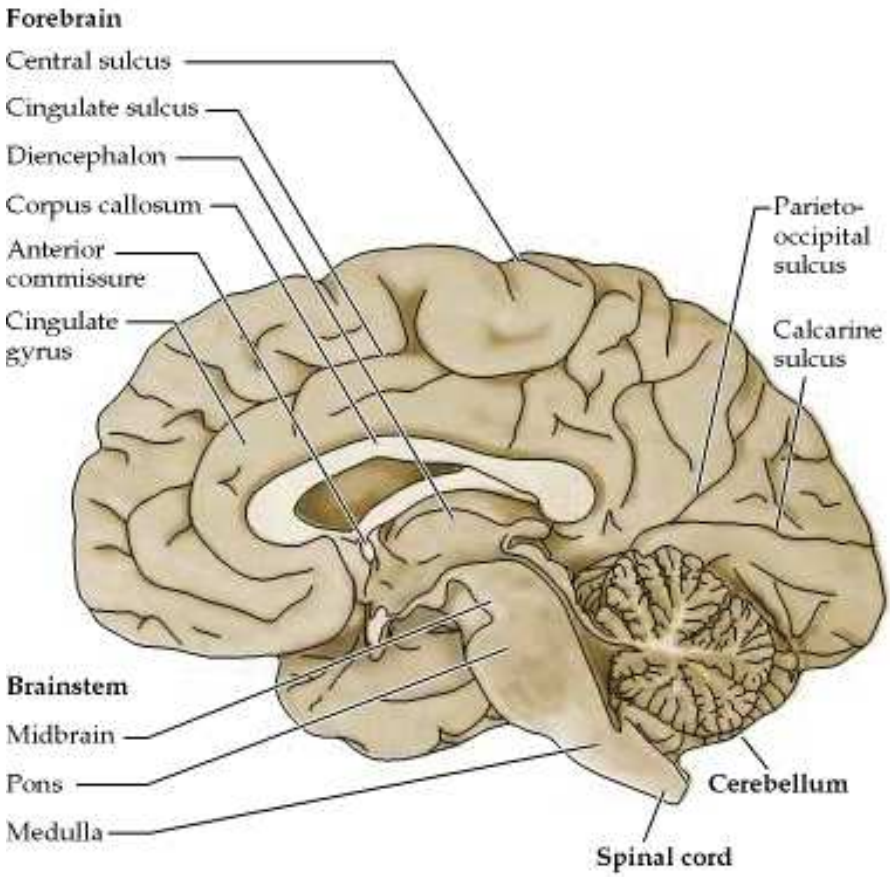

Figure 2.1: The medsagittal view of the brain and the brainstem (adapted from Purves et al., 2004)

The brainstem (see Figure 2.1) is considered to be structurally the most primitive part of the brain, but functionally it is the most vital, having control over such functions as breathing and conciousness. The physiology of the brainstem is as seen in Figure 2.2. Structures in the brainstem handle motor and sensory information coming from different modalities, including upstream pathways for sensory information and downstream for motor information.

The midbrain handles visuomotoric functions (in the superior colliculus, the oculomotor nucleus and the trochlear nucleus), visual reflexes (in the pretectal region), auditory relays (in the inferior colliculus) and motor coordination (in the red nucleus).

The pons relays connections from the cortex to the spinal cord, brainstem and cerebellar regions. Incoming information from the auditory and vestibular modalities have primary

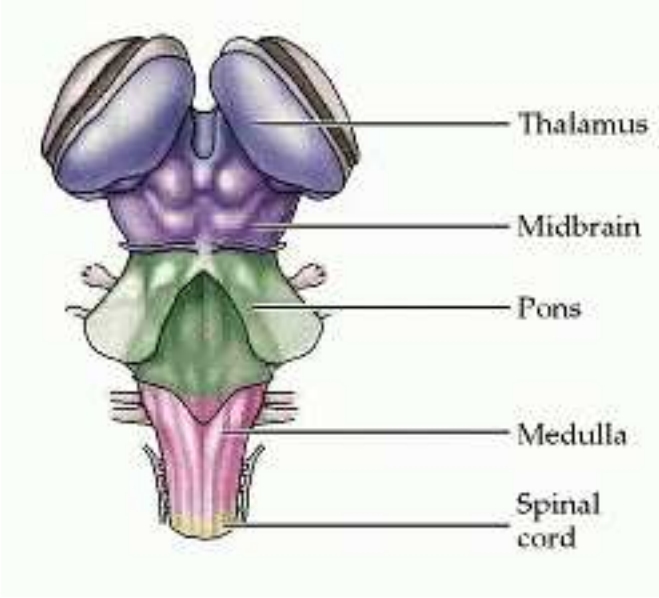

Figure 2.2: Anatomy of the human brainstem (adapted from Purves et al., 2004)

CNS synapses in the pons and neurons from the face and mouth areas deliver their information to the sensory and motor nuclei.

The medulla has nuclei for ascending somatosensory information from the spinal cord up to the thalamus and the somatosensory cortex and corticospinal ascending connections of motor control to the spinal cord. There is also nuclei for relaying information from the cortex to the cerebellum and sensory nuclei handling information from the facial area.

The cerebrum

The cerebrum is divided into left and right lobe by the interhemispheric fissure. Axons running through the corpus callosum create connections between the two lobes of the brain. Connections to and from different modalitites to the cerebrum are usually crossed, left side of the body connecting to the right lobe and vice versa.

The cerebrum has physiological landmarks based on the folding of the gray matter. The most prominent features are the central sulcus, which marks the separation of the frontal lobe from the parietal lobe, the lateral fissure, which in turn separates the temporal lobe from the frontal and parietal lobes, and the parieto-occipital sulcus and the pre-occipital

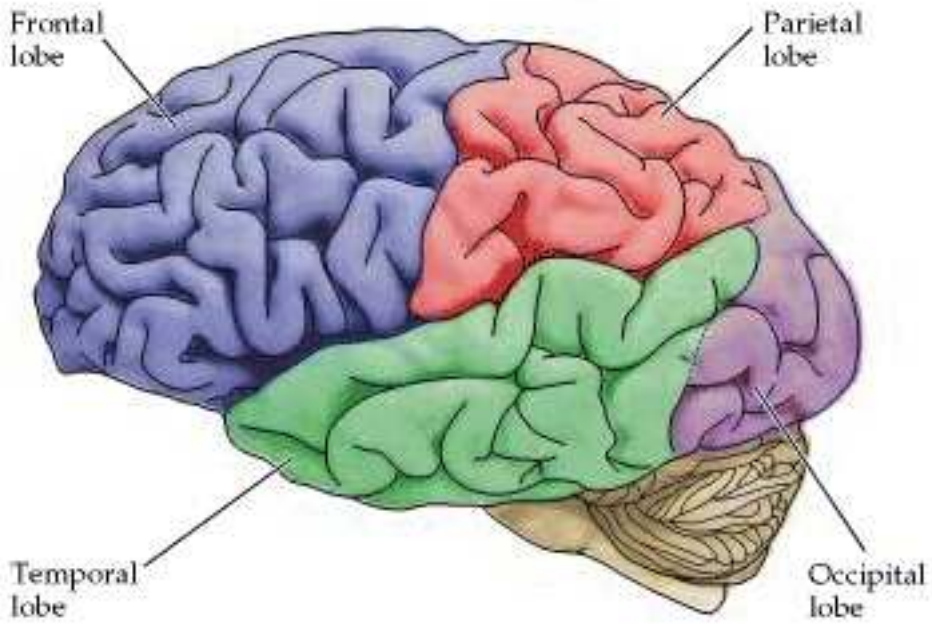

Figure 2.3: Lobes of the cerebrum (adapted from Purves et al., 2004)

notch, separating the occipital lobe from the parietal and temporal lobes.

Rough division of functionality between the lobes of the cortex is as follows: the frontal lobe is involved mainly in motoric functions; the parietal lobe in somatosensory functions; visual processing is assigned to the occipital lobe and auditory processing to the temporal lobe. Associations between different modalities are made on many areas of the cortex with no single designated area.

In 1909 German neuroanatomist Korbinian Brodmann used tissue staining methods to determine fifty-two areas that differed from each other based on cellular structure and organization. These divisions have later been found to correspond to the functional differentation over the cortex, with some refinement and subdivision. For example the primary visual area (V1) constitutes of the Brodmann area 17 and the primary auditory area of the Brodmann areas 41 and 42.

In the scope of this thesis the focus is not on the areas of the cortex, but on the brainstem.

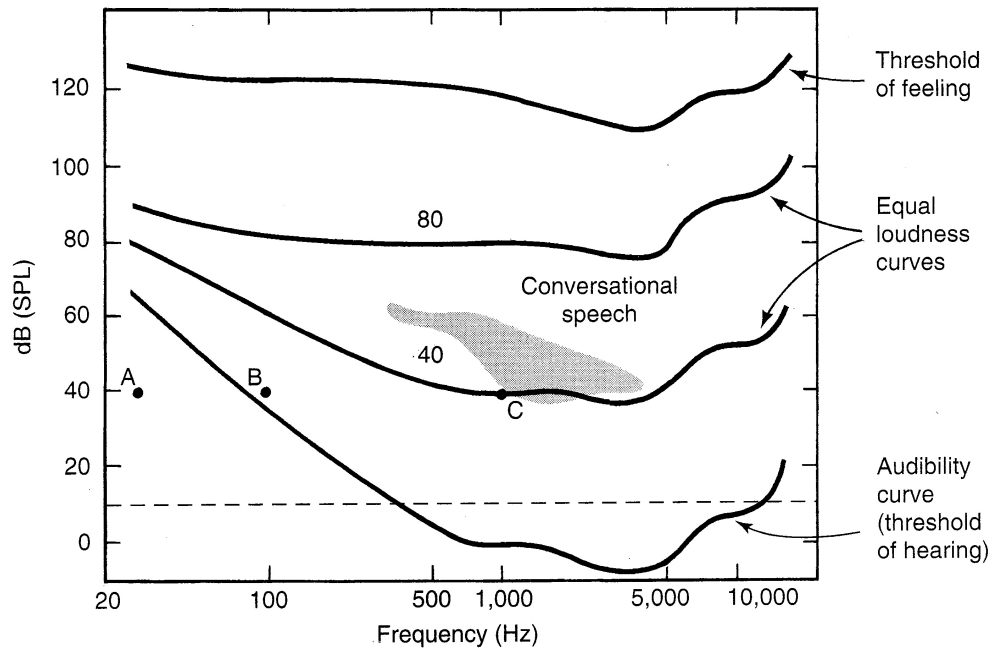

Figure 2.4: Equal loudness curves and the range of hearing (adapted from Goldstein, 2002)

2.1.2 Human ear and hearing

Sound transmission over a medium is based on the medium's particles' ability for having elasticity and inertia. Air as a medium is consisted of randomly moving molecules which a physical vibration from a sound source pushes into synchronized motion, causing variation to the otherwise static air pressure. More in depth explanation about the transmission of sound can be found from Yost (2000, Chapter 3).

Human has a binaural hearing system: two ears on both sides of the head enables accurate distinction of the direction of the heard sound and makes the human auditory system less vulnerable to one-sided damage (Karjalainen, 1999).

The range of human hearing falls between frequencies 20–20 000 Hz. (Yost, 2000; Goldstein, 2002). Sound pressure level (SPL) is a way to express the intensity of a sound in respect to a sound pressure of 20 μPa, determined to be the smallest amount of air pressure change needed for a person with a normal hearing level to hear a sinusoidial sound between 1000–4000 Hz (Yost, 2000). Depending on the SPL of the sound, a sound with

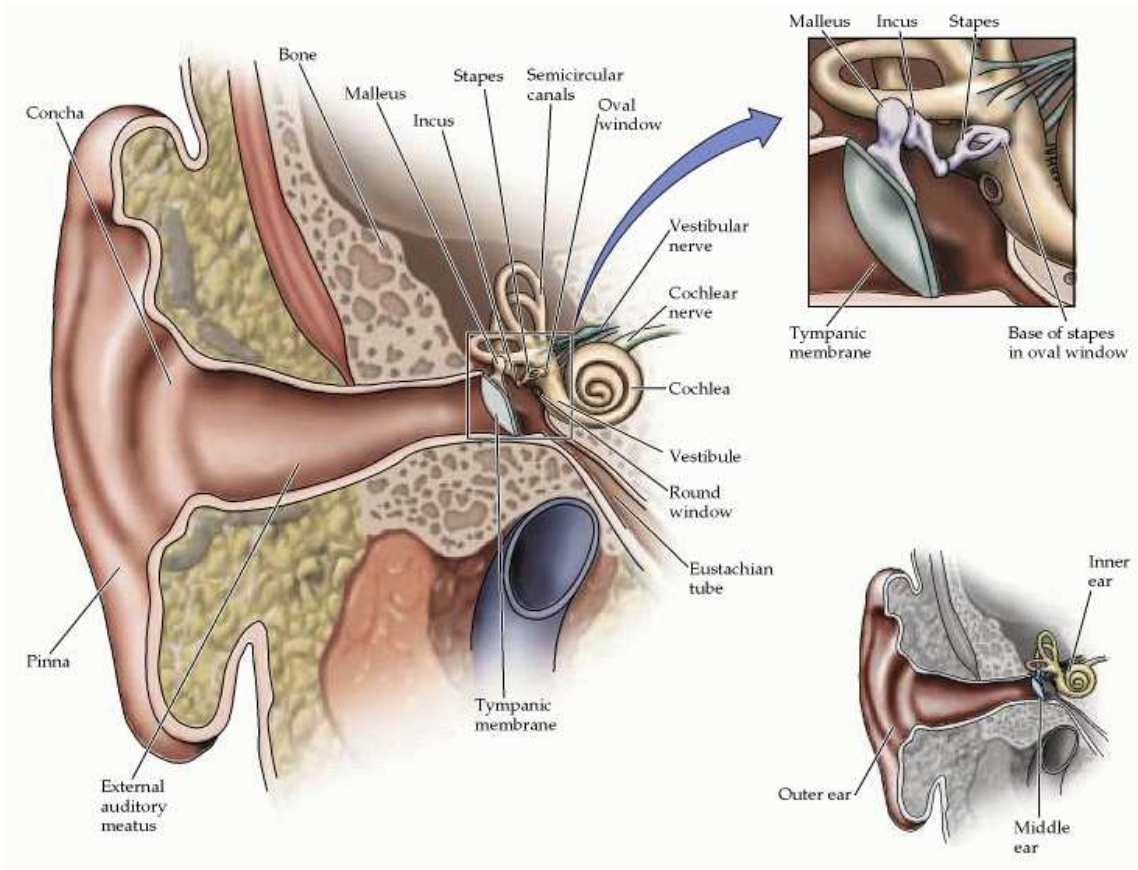

Figure 2.5: Anatomy of the human ear (adapted from Purves et al., 2004)

identical frequency can be perceived as being more loud than a sound with lower SPL, as can be seen in the equal loudness curves in Figure 2.4.

The ear consists of three parts. Outer ear consists of the pinna, external auditory canal and the tympanic membrane, commonly known as the eardrum. Middle ear includes the auditory ossicles: malleus, incus and stapes; the oval window and the Eustachian tube. Inner ear includes the cochlea and the semicircular canals (vestibular system). (Karjalainen, 1999; Yost, 2000)

Structure of the outer ear

The outer ear is a passive element in the auditory system with linear filtering function defined purely by physical and acoustic laws. Changes in physical structure of the pinna – such as trauma – would affect directional and spatial localization of sound, but not

the neurological auditory system (Karjalainen, 1999). The external auditory canal has a resonance frequency of about 2.5 kHz (Yost, 2000) and thus causes amplification to the frequency range of human speech at 2–5 kHz (Goldstein, 2002). Auditory canal provides protection to the delicate structures deeper in the ear against any foreign objects and acts as a regulator of humidity and temperature (Yost, 2000; Goldstein, 2002).

Outer ear and middle ear are separated by the tympanic membrane which functions as a sound transmitter into the middle ear as pressure changes cause the tympanic membrane to vibrate.

Structure of the middle ear

Vibrations of the tympanic membrane are amplified in the ossicles. Main portion of the amplification is due to the size difference between the originating tympanic membrane (approx. 55 mm^2) and the receiving oval window (approx. 3.2 mm^2). In addition, some amplification occurs in the lever-like joins between the ossicles, adding up to a total theoretical gain of 33 dB (Yost, 2000). Amplification is necessary because of the medium change from air to liquid which occurs when entering the inner ear. Due to the impedance mismatch of the two medium, airpressure alone would not be effective enough in transmitting sound waves to the inner ear (Yost, 2000; Goldstein, 2002).

The middle ear protects the auditory system from damage in two different ways. The small muscles attached to the ossicles can contract due to high intensity sound, thus stiffening the ossicle chain and reducing the intensity of the transmitted pressure into the inner ear. This reflex is stronger at frequencies below 2 kHz, helping speech perception in noisy environments (Bear et al., 2001; Yost, 2000). Against sudden loud noises the attenuation reflex is too slow with 50–100 ms delay, but it is effective against long lasting high-intensity sounds. Other proposed functions for these muscles include the suppression of ones own voice during speech and possibly increasing the dynamic range of hearing (Bear et al., 2001).

As the middle ear is an air filled cavity separated from the outer ear by the flexible tym-

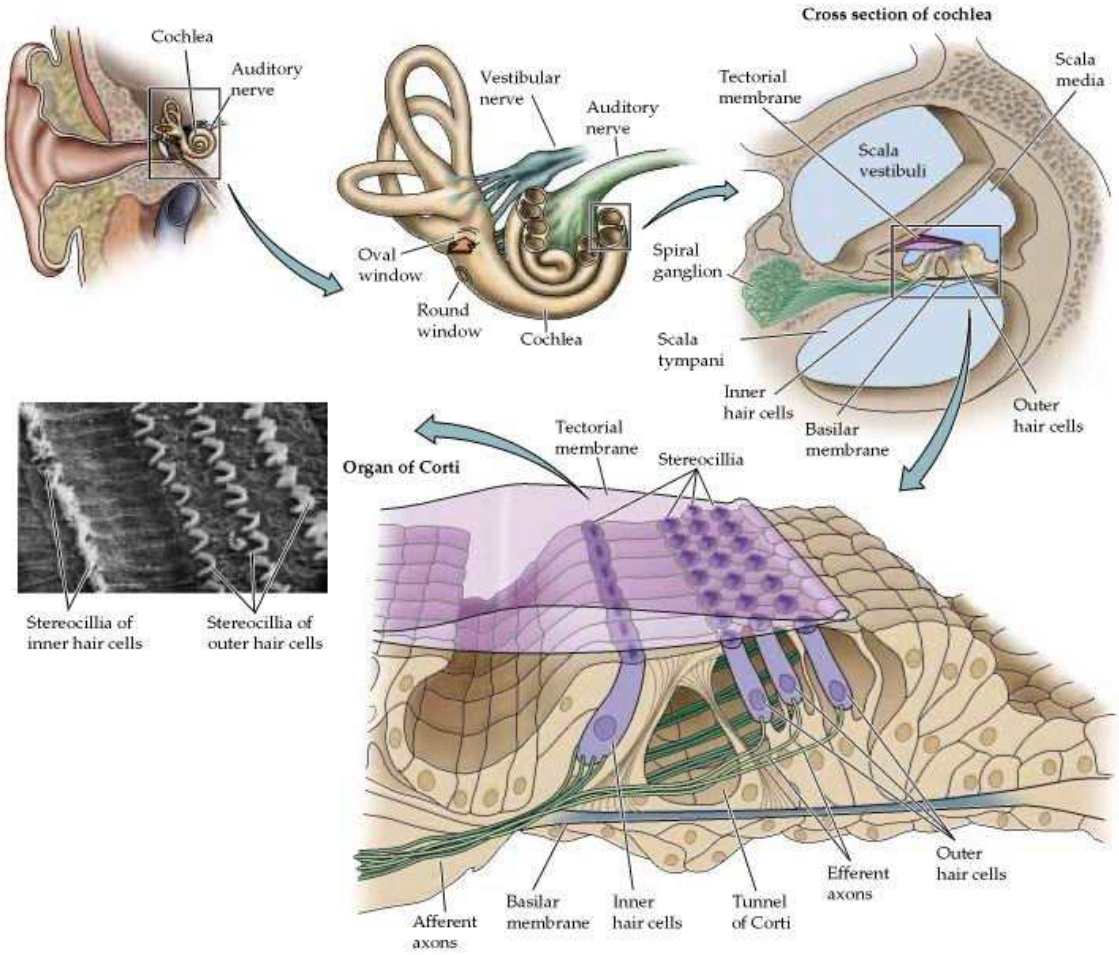

Figure 2.6: Anatomy of the inner ear and cross section of the cochlea (adapted from Purves et al., 2004)

panic membrane, large changes in surrounding airpressure outside – for example when flying – could cause the tympanic membrane and ligaments supporting it to stretch, causing pain and possible damage to the structures. The eustachian tube, which is covered by a valve-like structure, connects the middle ear to the nasal cavity. Yawning or swallowing opens the valve and thus equalizes the pressure inside the middle ear to that of the surrounding atmosphere, preventing damage (Bear et al., 2001; Yost, 2000; Goldstein, 2002).

Structure of the inner ear

The inner ear is further divided into three different sections: the semicircular canals, the vestibule and the cochlea. The semicircular canals connecting to the vestibule are mainly responsible for the sense of balance based on the movement of the liquid inside the three-dimensionally arranged canals (Yost, 2000).

The cochlea is a bony tube of approx. 2 mm in diameter and 32 mm in length, narrowing towards its end (the apex or the helicotrema) and coiling around its center (the modiolus) 2.5 times into a shell-like form. Inside the cochlea are three sections: scala vestibuli spanning from the oval window and scala tympani spanning from the round window, and they are joined together at the helicotrema. Between these passages lies the scala media (or the cochlear partition) which contains the actual structures for transforming sound waves into electrical signals. The scala media is lined by Reissner's membrane, the basilar membrane and the stria vascularis, forming a closed partition separate from the scala vestibuli and the scala tympani (Yost, 2000; Goldstein, 2002).

The basilar membrane widens from the oval window towards the helicotrema, being stiffer at the beginning and more flaccid at the apical end of the cochlea. This causes changes in the vibratory pattern along the basilar membrane, creating the basis for the frequency differentiation in the cochlea. High frequency sounds cause vibrations in the beginning of the basilar membrane while the low frequency sounds travel further towards the apex.

The whole of the cochlea is filled with fluid, the scala vestibuli and the scala tympani containing perilymph and the scala media containing endolymph. The main physiological difference between the two is the K^+ to Na^+ ratio: perilymph is like the cerebrospinal fluid with concentrations of 140 mM Na^+ and 7 mM K^+ and endolymph resembles intracellular fluids with concentrations of 1 mM Na^+ and 150 mM K^+ (Bear et al., 2001; Yost, 2000). This creates the highest potential difference inside the human body, 80 mV, which is maintained by active ionic transport process at the stria vascularis.

The stapes conduct the mechanical movement through the oval window to the perilymph of scala vestibuli, causing the fluid to move. This movement is continued through the

helicotrema into the scala tympani and to the round window.

Inside scala media and on top of the basilar membrane lies the organ of Corti, which is covered by the tectorial membrane. The organ of Corti includes two types of hair cells, inner and outer hair cells, and various other types of supporting cell structures. The hair cells have fine hairs (cilia) on their upper surface. The pressure changes in the fluid cause the scala media to move in an up-down motion, moving the basilar membrane and the hair cells upon it as well. Due to the different pivoting point of the basilar membrane in contrast to that of the tectorial membrane, as the basilar membrane and hair cells move upwards, the tectorial membrane moves forwards, thus bending the cilia away from its pivoting point. Downwards moving basilar membrane and hair cells cause the tectorial membrane to bend the cilia back towards its pivoting point (Yost, 2000). The movement of the cilia causes the corresponding hair cell to either de- or hyperpolarize, depending on the direction (Bear et al., 2001). As the basilar membrane vibrates with different frequencies along its length, this causes the cilia to send onwards singals that are already separated into frequency ranges. This tonotopic arrangement is preserved all the way to the primary auditory cortex (see Figure 2.7) (Purves et al., 2004).

Hair cells form connections with neurons residing in the spiral ganglion within the modiolus. These neurons send their axons to form the auditory nerve, also known as the VIII cranial nerve. The auditory pathway (see Figure 2.8) goes to the superior olivary complex in the medulla and lateral lemniscus, where sound localization is determined on binaural cues. The inferior collicus in the midbrain carries out frequency processing and space localization. The medial geniculate nucleus in the thalamus relays the signals to the primary auditory cortex in the superior temporal gyrus. (Purves et al., 2004)

Auditory Brainstem Response

This section is mainly based on the publications of Niedermeyer and da Silva (1999) and Jacobson (1985).

Auditory brainstem response (ABR) describes the first milliseconds of neural activity in

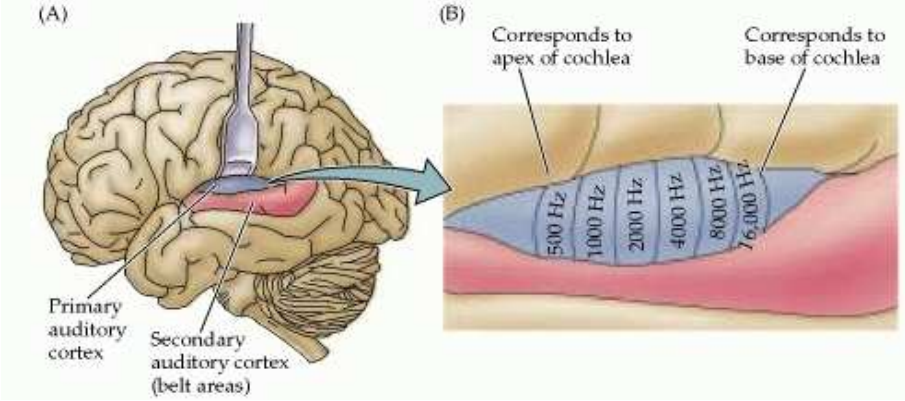

Figure 2.7: The tonotopic organization of the primary auditory cortex (adapted from Purves et al., 2004)

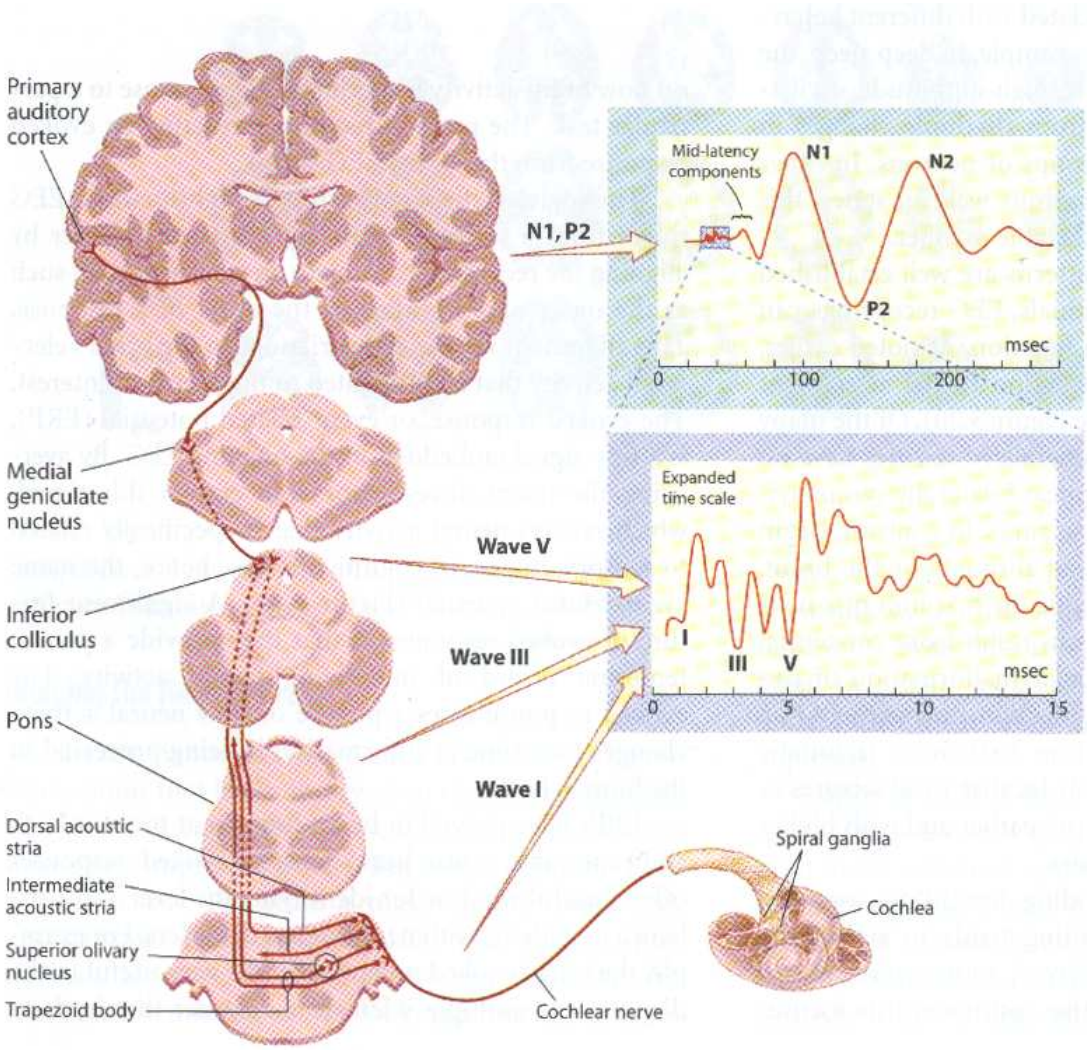

Figure 2.8: The human auditory pathway and the generation sites of the different parts of the auditory brainstem response (adapted from Gazzaniga et al., 2002)

the human auditory system after an auditory stimuli reaches the ear. The earliest stage represents firing patterns in the cochlea and the eighth cranial nerve.

The evoked potentials caused by an auditory stimulus can be recorded either from the scalp, from inside the ear canal or invasively from the nerve structures. Scalp recordings allow the recording of overall activity over the cortex. The downside is the lower signal to noise ratio if focusing only on the auditory brainstem response.

Ear canal recordings are done by inserting a thin electrode tip into the ear close to the eardrum and thus getting closer to the actual generation sites of the earliest signals. This setup also allows the capture of the cochlear michrophonic – the actual electrical signal emitted by the cochlea in response to the stimulus.

ABR is used clinically in determining possible neural deficiencies in the auditory nervous system. This is possible because the normal waveform is very well documented. When using the Jewett nomenclature to describe the ABR (Figure 2.9), the first five wave peaks are labeled with Roman numerals and can be detected from averaged responses by trained eye. Latency changes between peaks, depleted amplitudes or missing peaks can reveal underlying neural problems. ABR can also be used in determining auditory sensitivity in subjects because there is a strong correlation between stimulus intensity and waveform. In psychophysical experiments, auditory brainstem response measurements combined with performance evaluations from different tasks can be used to study for example learning difficulties in children and adults (Johnson et al., 2007; Cunningham et al., 2001).

2.2 Measuring activity in the brain

2.2.1 Electroencephalography

This section is mainly based on publications by Malmivuo and Plonsey (1995), Niedermeyer and da Silva (1999) and Bear et al. (2001).

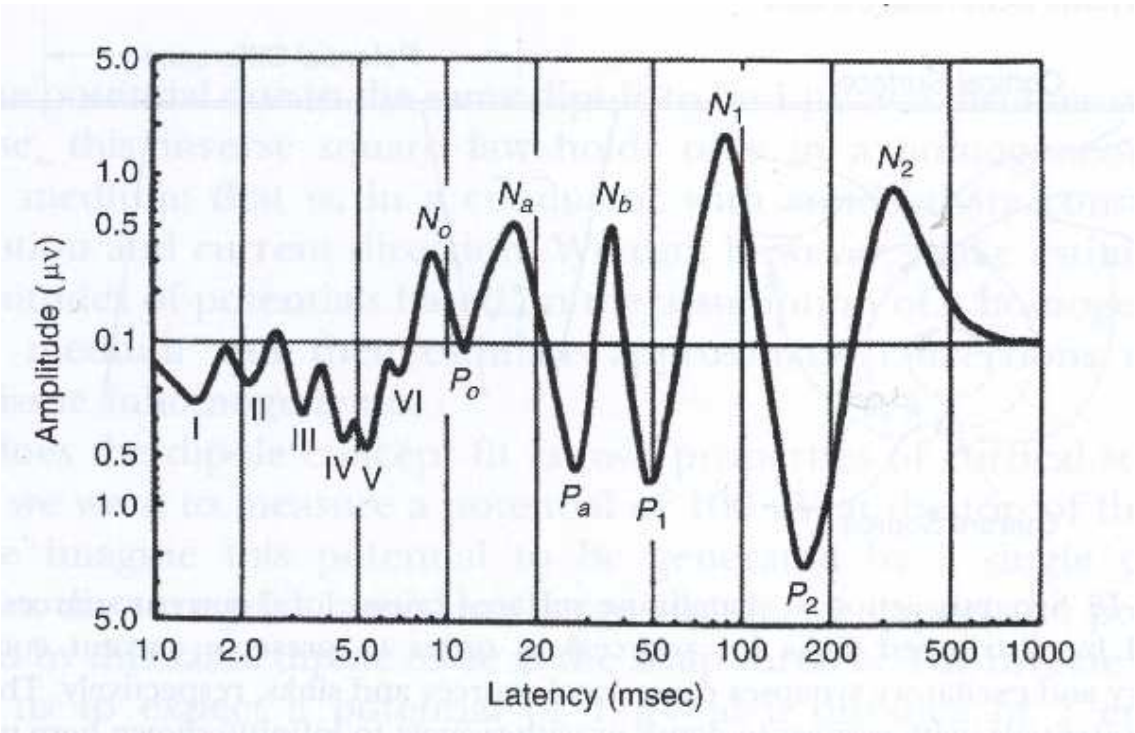

Figure 2.9: Wave model of the auditory brainstem response (adapted from Nunez, 2006)

The primary concept of the electroencephalography is to present the electrical activity measured from the brain in respect to time. The visual representation of these electrical potentials is called electroencephalogram. The first reported EEG-recording from the human brain was conducted by Hans Berger in 1924 (Malmivuo and Plonsey, 1995; Niedermeyer and da Silva, 1999). From his method of placing electrodes to the front and the back of the head, EEG-recording has come a long away incorporating invasive and non-invasive techniques and a multitude of electrode placements. In the early days the use of EEG was limited only to observing changes in the continuous spontaneus activity, which has a large amplitude in contrast to the background noise and can be visually observed from the EEG-scroll. In 1947 George D. Dawson introduced a new method of summation by superimposition for detecting evoked potentials which have very low amplitude compared to other EEG activity and noise, allowing the observation of neural responses to different types of stimuli (audio, visual etc.). Using repeating stimuli it is possible to attain repeating neural responses and when these responses are summed together, the similar firing patterns create distinctive waves that stand out from the background noise, which in turn is diminished during the summation process because of its random nature

(Malmivuo and Plonsey, 1995; Niedermeyer and da Silva, 1999).

EEG monitoring is used for both clinical and scientific research purposes: in medical situations the EEG activity may reveal known patterns corresponding to underlying diseases. Another example of medical use are brain-computer interfaces corporating continuous EEG monitoring have been developed to help patients suffering from movement disabilities (Kauhanen et al., 2007). In recordings with scientific goals - such as described in this thesis - correlation between the subject's actions or state of conciousness and patterns found in the EEG are explored after incorporating various statistical analyzing methods to the raw EEG signal.

2.2.2 Electrode placement

Because the measured electrical current on the scalp is the sum of signals originating in multiple generation sites and the accurate place of origin cannot be determined (the inverse-problem, Malmivuo and Plonsey (1995)), in order to get generalizable and comparable results, standardized electrode placement schemes have been developed. The scheme used in this study is the extended international 10-20 system with 32 electrode positions standardized by the American Electroencephalographic Society.

The 10-20 system uses percentage measurements (Figure 2.10) between electrodes starting from two reference point: nasion, the depression where nose meets the forehead and inion, the lowest point of the skull midline at the back of the head where a prominent bump can usually be seen.

2.2.3 Recording methods

In non-dreaming sleep or in comatose state the spontaneous, self-sustaining activity of neurons is highly synchronized. This results in high amplitudes on a low frequency range. When information is being processed on the cortex, be it in a dreaming sleep or awake, the activity is no longer clearly synchronical due to multiple generation sites. The cognitive

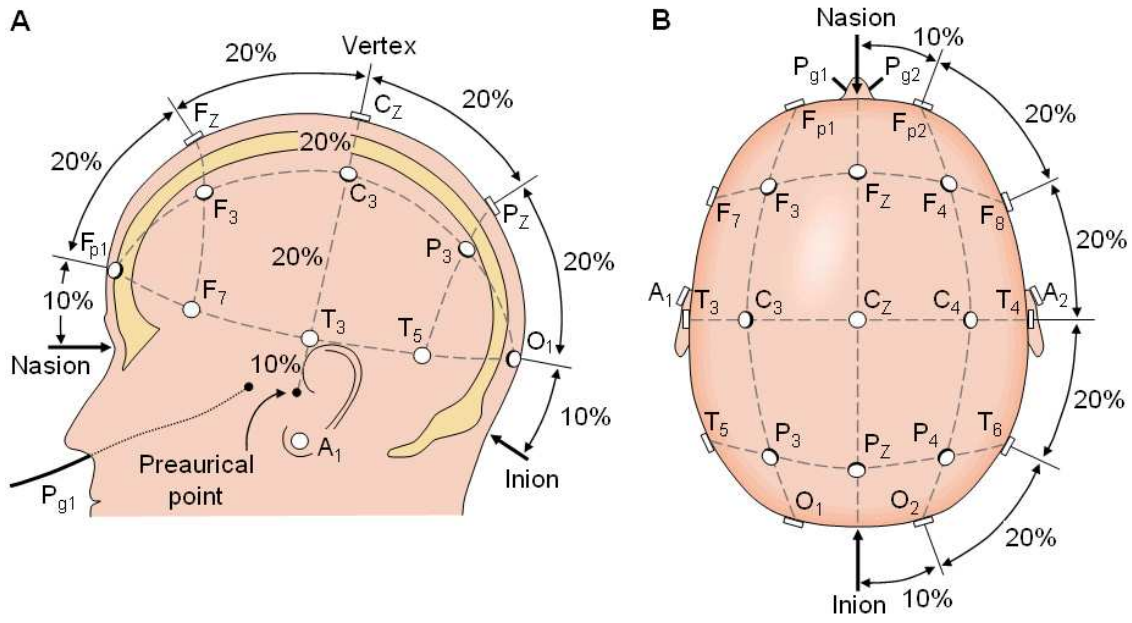

Figure 2.10: Percentage measurement principals of the international 10-20 electrode montage system (adapted from Malmivuo and Plonsey (1995))

functions are located in different areas of the cortex, resulting in much lower amplitude scale.

The amplitude of the EEG-signal is low in retrospect to surrounding noise amplitudes. While invasive methods produce amplitudes in the range of 1–2 mV, non-invasive scalp measurements produce amplitudes in the range of only 100 μV. (Malmivuo and Plonsey, 1995)

The metal electrodes used in aquiring the EEG-signal can be placed either inside the skull within the brain or non-invasively on the scalp. Invasive methods can produce spatially more accurate results but are not that common in basic research such as this thesis. Modern day use of invasive methods incorporate monitoring during brain surgery and brain-computer interface applications for disabled or locked in state patients (Hochberg et al., 2006).

Electrode montages

External scalp electrodes can be attached to the skin one by one or, as in this study, using a specially designed cap which has set places for the electrodes. Using an elastic cap has the advantage of being faster than attaching single electrodes one by one to the skin. The biggest problem with using such cap is the varying head size of the subjects - to achieve perfect fit, multiple sized caps should be used. If the subject has a significantly different head size than what the cap is designed for, the electrode montage is no longer valid.

2.3 Integration of visual and auditory speech - literature review

Multiple researchers have found evidence of visual speech affecting the processing of auditory pathway signals in higher areas of the brain. Effects on the N100 and P300 cortical steady state responses are widely researched and confirmed.

2.3.1 Audiovisual integration

A commonly used method in estimating if integration of different modalities has occured, is the additive model. When measuring electrical fields such as in EEG, fields add up to each other linearly. In this manner it is possible to compare ERP responses to unimodal stimuli to ERP responses to multimodal stimuli. If no integration between different unimodal processes ocurr, the sum of the unimodal responses will equal that of the multimodal response. In audiovisual cases, this effect can be postulated as AV-(A+V). If the difference between the bimodal and the unimodal counterparts is not zero, some level of integration has occurred. (Besle et al., 2009)

Visual and auditory speech integration at cortical level

The most known visual and auditory speech integration example is likely to be the McGurk-effect (McGurk and MacDonald, 1976). The original experiment used only behavioral testing, but as the effect has later been found out to be both reproducable and non-habitable, it has been used in conjunction with EEG and brain imaging methods to research the underlying neural mechanisms.

fMRI-based studies

Calvert et al. (1997) found in their fMRI study that the same areas in the primary auditory cortex and auditory association cortex located in the lateral temporal auditory cortex (Brodmann areas 41, 42 and 22) were activated not only by auditory stimuli consisting of words, but also by viewing a video presentation of a person silently mouthing numbers without any auditory stimuli. Additionally they found similar activation when the subjects were shown visual pseudospeech (speechlike nonsense). However, these activations were not present when the subjects viewed non-linguistic facial movements. Calvert et al. (1997) postulated that this result provided physiological base for the McGurk effect by showing that lipreading affects auditory perception of speech at a pre-lexical level.

Pekkola et al. (2005) observed in their fMRI study activation in Heschl's gyrus and primary auditory cortex in subjects observing visual speech. The visual stimuli used in this study was nearly identical to the visual stimuli used in this thesis (same original video, differences in sequence order and length). Pekkola et al. (2005) used the expanding rings condition (or moving circles as described in the publication) as the control condition, resulting in significantly lower response signal levels when compared to the visual speech condition. Significant left hemisphere dominance was observed in the results, suggesting specialization in visual speech processing. The researchers suggested that possible explanations might be either converging input from visual modality to the auditory cortex; articulation movements enhancing primary auditory cortex activation due to learned connection between articulation and speech in reaction to the scanner noise, or subvocal-

ization or subversive speech which was not explicitly discouraged.

Electrophysiological studies (EEG/MEG)

Klucharev et al. (2003) implemented an ERP-based study of the possible integration of auditory and visual speech. Stimuli combinations consisted of congruent and incongruent audio-visual representations of Finnish vowels versus auditory only and visual only unisensory situations. As a fundamental theory they used the additive model of integration, in which integration is presumed to occur if the sum of the unisensory auditory (A) and unisensory visual (V) ERP magnitude is not equal to that of the audio-visual (AV) ERP.

Significant spatial and time-based differences were retrieved based on the ERP data. Klucharev et al. (2003) suggested two different types of integration: early latency non-phonetic integration, in which the congruent and incongruent AV ERPs did not differ significantly from each other, and later latency (from 150 ms onwards) phonetic integration, in which there was a significant difference between the congruent and incongruent AV situation ERPs. Non-phonetic integration was found to be lateralized to the right side of the brain, with suggested originating sites at the extrastriate visual cortex and non-primary auditory cortices. In the phonetic integration findings the incongruent AV elicited ERPs were found to be larger in magnitude in earlier latencies and only at the last point of significance at 325 ms did the congruent AV elicited ERP surpass that of the incongruent one. Origination sites to the phonetic activations were suggested to reside in the posterior temporal cortex (posterior part of STS) and parietal and inferior temporal regions.

van Wassenhove et al. (2005) used auditory and visual speech syllables (/ka/, /pa/, /ta/) to study the effects of visual speech to cortical N1 and P2 auditory ERP responses. They found suppression of the N1 and P2 responses in audiovidual speech conditions when compared to the audio-only condition. In addition they reported significantly different latencies of N1 and P2 between the different syllables, dependending on how well the subjects identified the correct syllable in the visual-only condition.

Besle et al. (2004) used auditory and visual unimodal and congruent audiovisual stimuli in their EEG experiment. They found that subjects detected the audiovisual target stimuli faster than the unimodal stimuli. In their EEG-analysis they used the same additive model as Klucharev et al. (2003), resulting in supressed N1 activity in the auditory cortex in the audiovisual condition when compared to the sum of the unimodal activities ([A+V]).

Kauramäki et al. (2010) used similar visual stimuli as in this thesis (Finnish vowels) combined with pure tones to study the effects of lipreading and silent speech production to auditory cortex responses. They found that both observing visual speech and silently producing the same vowels suppressed the N100m response, with dominance in the left hemisphere, in comparison to the expanding rings condition. They suggest that this suppression is caused by an efference copy signal from the speech production system affecting the auditory processing on the cortex in a top-down manner.

Auditory brainstem response specific studies

The auditory brainstem response has long been used in clinical diagnosis of hearing problems. Cunningham et al. (2001) found that children with reading based learning problems with no hearing deficits showed longer wave V latencies and diminished spectral component magnitudes of FFR when listening to auditory stimuli in background noise as opposed to children showing normal learning curve.

Musacchia et al. (2006) measured the auditory brainstem responses to unimodal auditory and concordant and conflicting auditovisual conditions. The auditory stimulus used by the group was similar to the auditory stimuli used in this thesis (/da/, 100 ms in length with 10-ms consonant burst, 30-ms formant transition and 60-ms steady-state vowel). They found that the size of the initial 10 to 30 ms section of the ABR in both of the the audiovisual conditions were suppressed when compared to the audio only -condition. In addition they found statistically significant increased latency in the onnset portion of the responses in the audiovisual conditions, compared to the unimodal auditory stimuli response. Musacchia et al. (2006) suggest, based on their results, possibility of speech specific processing in the brainstem level triggered by articulatory gestures.

2.4 Aim of the present study

The aim of the present study is to (1) qualify if the EEG-equipment described in Section 3.5 is able to reliably record the ABR and (2) observe if attending to visual speech affects early auditory processing in the brainstem in comparison to spatially similar non-speech visual stimuli.

Chapter 3

Methods

This chapter describes the methods used in the aquisition of the data in this thesis.

3.1 Pilot study

A three subject pilot study was performed prior to the actual experiment. In this group the subjects underwent each condition only once and the results are thus not directly comparable with those of the later subjects. The pilot study data is not included in the results.

3.2 Subjects

Nine healthy, right-handed volunteers participated in the experiment. All subjects were native Finnish-speakers and reported normal hearing and normal (or corrected to normal) vision. Five of the subjects were female, four males. Mean age of the subjects was 25.2±s.d. 6.3 years. Subjects did not receive monetary compensation. Instead, four of the subjects participated in the experiment as a part of their Cognitive Neuroscience -course and received one (1) additional point as compensation to their final course evaluation. The

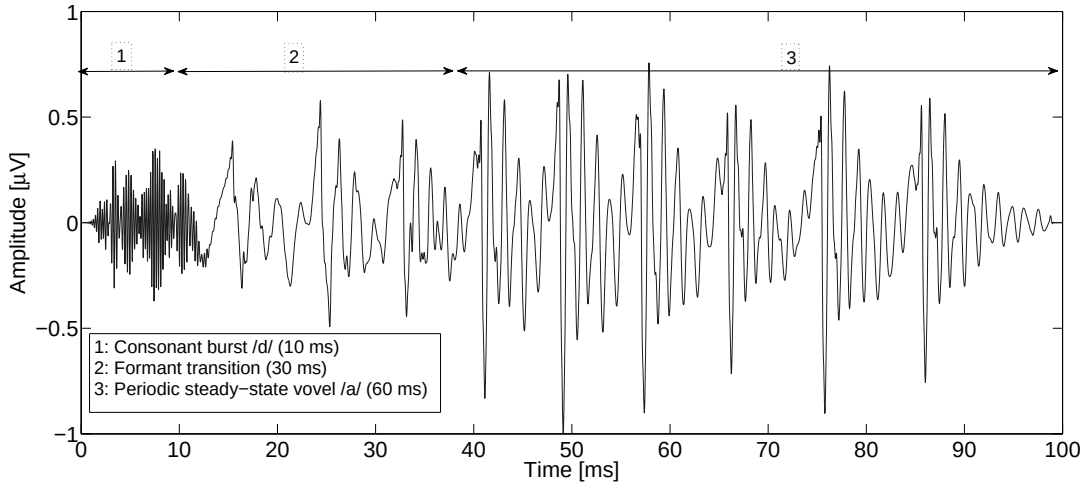

Figure 3.1: The auditory stimuli: 100 ms /da/ syllable

course personnel was not involved in this study and only received the student numbers of those that participated in the experiment. A written informed consent was obtained from each subject prior to the experiment. The data of two subjects had to be omitted from the final results due to technical issues. The analyzed data is thus consisted of seven subjects (mean age 26.4±s.d. 6.8 years), four females and three males.

3.3 Stimuli

The auditory stimuli used was a synthesized /da/ syllable of 100 ms in length. It consisted of 10 ms consonant burst, followed by 30 ms formant transition and 60 ms periodic, steady-state vowel (see Figure 3.1). The auditory stimuli was the same as Musacchia et al. (2006) used in their auditory brainstem study. The auditory stimuli was presented with 151 ms onset-to-onset ISI, resulting in repetition rate of 6.62 / s. Alternating polarities of the same waveform were used in order to avoid the possible signal contamination from the cochlear michrophonic response (see Section 2.1.2). The sound file used in the experiments was edited to include 30-ms silent period before sound onset to provide a baseline period for the EEG recording.

The visual stimuli consisted of predefined and pre-edited video sequence depicting a fe-

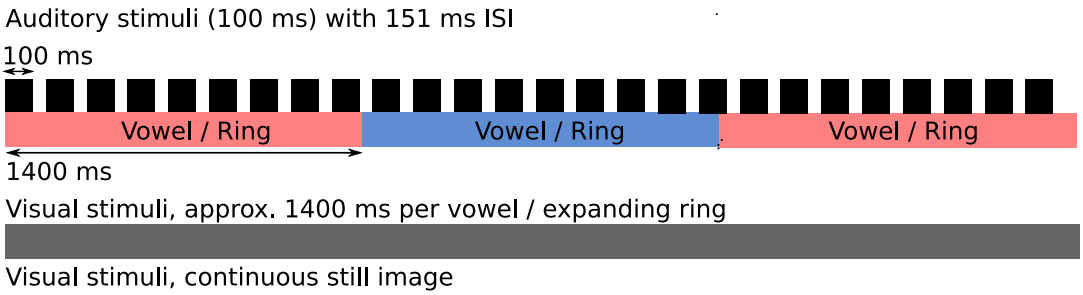

Figure 3.2: Example of the stimuli distribution during an experiment

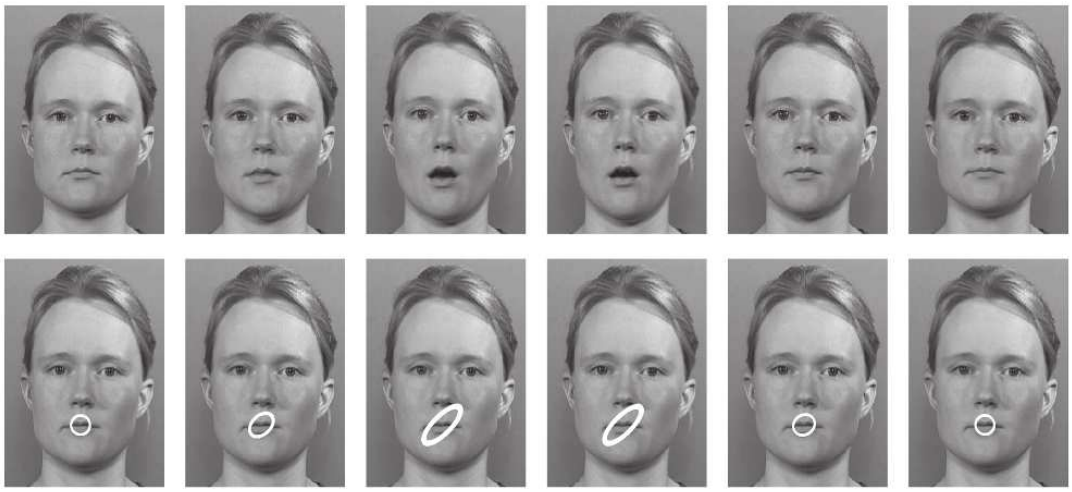

Figure 3.3: Example of the visual stimuli /a/ and the expanding ring (adapted from Kauramäki et al., 2010)

male face in three different conditions. Similar visual stimuli have been used in studies by Pekkola et al. (2005) and Kauramäki et al. (2010). The still images were aquired from video footage commonly used in our laboratory, presenting a female actress mouthing vowels. In the still condition there was no induced movement as only one frame was presented - a neutral facial expression without any emotions or visual speech cues. In the vowels condition the video sequence was edited so that the face showed articulatory gestures of Finnish vowels /a/, /i/, /o/ and /y/. In the third condition the sequence consisted of expanding blue rings / ovals superimposed over the mouth region of the neutral still face, thus creating temporally and spatially similar movement perception as the vowel condition, but without the linguistic content. Figure 3.3 shows an example of both types of stimuli, vowel and expanding ring.

3.4 Experiment setup

The experiment room was an electrically shielded booth, isolated from external sound sources. The walls and ceiling of the booth were padded with sound proofing material, lowering the background noise level in the booth to 23 dB SPL(A) (for comparison, the background noise level in the adjacent EEG-observation room was 46 dB SPL(A) with 3 computers in the immediate proximity). The subjects were seated in a comfortable recliner chair in front of a 19 inch CRT display with an average viewing distance of 1.25 meters - subjects were not immobilized during the experiment, causing possible deviations in the distance. The auditory stimuli was presented through an Edirol MA-15D loudspeaker (Roland Systems Group, Bellingham, WA, USA) located above the display, resulting in 63 dB SPL(A) sound level at the point where the subject's head was situated.

A computer running Presentation software version 9.81 (Neurobehavioral Systems, Inc., Albany, CA, USA), which was used to control the stimuli presentation, was located outside the shielded experiment booth. This was also the case for the computer running the BrainVision Recorder software (Brain Products GmbH, Gilching, Germany) used to record the EEG during the experiment. Connections between the computers and the equipment inside the booth were run through a shielded inlet into the room.

The hearing levels of the subjects were tested prior to the experiment and attenuation of the auditory stimuli set for each individual to achieve equal conditions for testing.

3.4.1 Task conditions

The experiment consisted of three conditions with visual and auditory stimuli. The conditions were named based on the visual stimulus presented and are hence forth referred as such: condition one as expanding rings, condition two as still face and condition three as vowels.

The auditory stimuli was the same over all conditions. The synthesized /da/-syllable (Figure 3.1) was repeated 3400 times in alternating polarities (1700 of both signal polarities),

resulting in total task length of 513.4 seconds (approximately 8.5 minutes).

In the expanding rings condition the subject viewed a still face of the female speaker superimposed with an expanding blue oval with the auditory stimulus playing in the background. The subjects were instructed to focus on the mouth region and indicate when two consecutive ovals expanded to the same direction.

In the still face condition the subject viewed a static image of the speaker's face showing on the screen while the auditory stimulus was presented. The subject was instructed to focus their gaze to the mouth area of the still face, in accordance to the other two conditions. The subject was not given any other specific task.

In the vowels condition the subject viewed preset visual presentation of the same female speaker mouthing silent Finnish vowels, while the auditory stimulus was presented. Subjects were instructed to focus on the mouth area of the face and were also given a task of following the vowels and indicating when two consecutive vowels were the same.

The subjects were given to hold an indicating device during the experiments, and they were instructed to give their answers by lifting their finger, which caused a trigger signal to be transmitted to the recording computer. These tasks were designed to keep the subject vigilant during the experiment, and to focus their attention to the correct area of the visual stimuli. Approximately 10 % of the visual stimuli were target stimuli.

3.5 EEG recording

A 32-channel actiCAP (Brain Products GmbH, Gilching, Germany) with the standard 10-20 system montage with re-attachable, active electrodes was used in the experiment. The cap removed the necessity of using glue when placing the electrodes. The conductive paste was applied with a syringe and blunt needle through a small hole on the electrode's plastic encasing. The paste used in the experiment was Electro-gel (by Electro-Cap International, Inc., Eaton, Ohio, USA). The 30 EEG channels used were Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10,

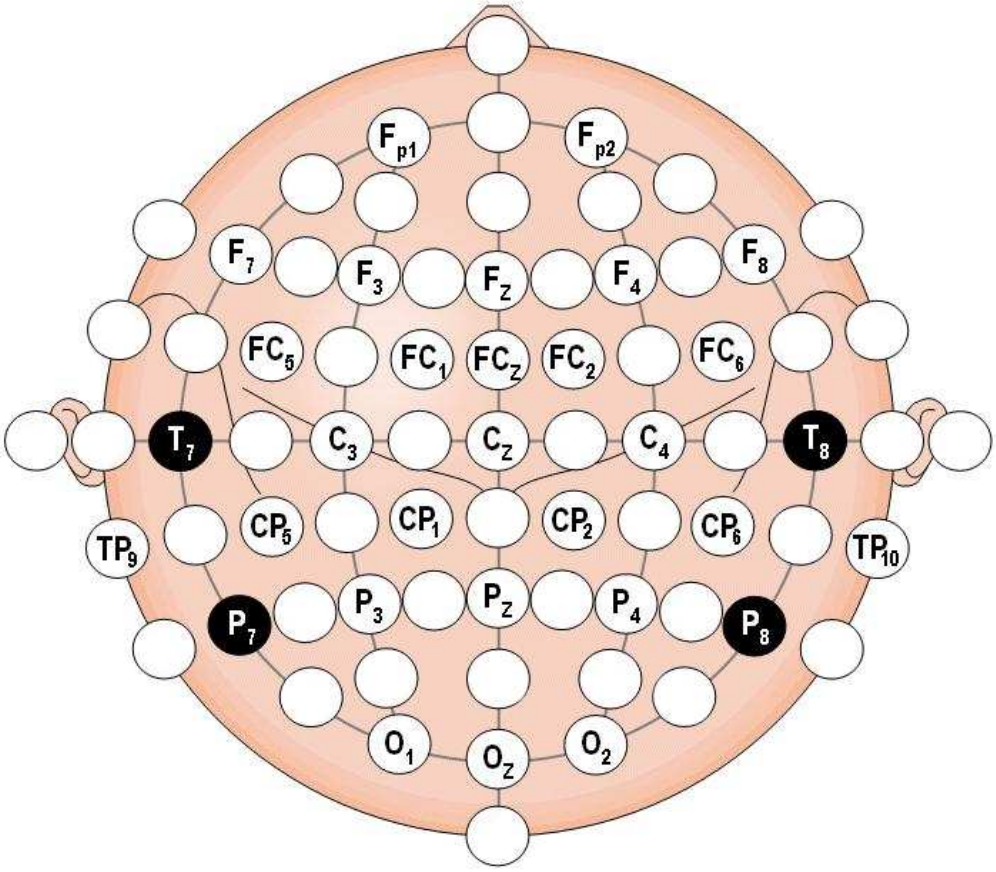

Figure 3.4: The electrode positions used in this experiment according to the international 10-20 system (adapted from Malmivuo and Plonsey, 1995)

P7, P3, Pz, P4, P8, O1, Oz and O2 (see placement of the electrodes in Figure 3.4). The reference electrode used was FCz. Two additional electrodes were used to monitor eye movement (EoC) and they were attached below and to the right side of the subject's right eye with two-sided skin tape intended for this use.

Brainvision Professional BrainAmp (by Brain Products GmbH, Gilching, Germany) was used in signal aquiring, incorporating 0.1–1000 Hz frequency band, 16 bit AD conversion and 5000 samples/channel/s. The electrodes were connected to the amplifier by the provided actiCAP connector box with a USB-connector. The actiCAP provides a program to monitor the impedance levels of the electrodes, and all channels were monitored prior to recording and in the middle of the experiment to make sure the impedance levels were below 5 kOmh.

Prior to the actual task runs, the subjects were shown a short introduction sequence (5 minutes) with the visual stimuli from the experiment. During this introduction, the levels of the EEG data and impedance levels were observed to find possible problems before the actual experiment. After this introductory sequence the actual task conditions were run in randomized order between subjects. Before each 8.5 minute task the subject was explained the upcoming task and what kind of stimuli he/she would be observing. Subjects were also offered the chance to have a short pause between the task runs. Subjects performed each task twice in randomized order.

The live EEG data received via the actiCAP interface was monitored and recorded during the experiment. Each condition was recorded as its own dataset. All 30 EEG data channels were recorded, as well as the two EoC-channels, and in addition to this, there was one extra channel fed by a recording microphone set in the experiment room to monitor the auditory signal and to offer a reference point to the actual raw EEG data. All in all, 33 channels were recorded.

The recorded data was initially analyzed offline with BrainVision Analyzer 1.0 software (by Brain Products GmbH, Gilching, Germany) with additional analysis (SNR, FFR, RMA, statistical analysis) done in Matlab (version R2007b, by The MathWorks Inc., Natick, Massachusetts, USA).

3.6 Data Analysis

3.6.1 Behavioral data

In the vowels and expanding rings conditions the trigger signals recorded from the subjects reacting to the target stimuli were analyzed and compared to the correct target stimuli timecode sequence. Response was judged to be a miss if the reaction occurred over 2000 ms after the target stimuli. False alarms (reactions to non-existing targets) were also taken into account. Hit rates, reaction times and d-prime values were calculated over the conditions. OpenOffice.org Calc (version 3.1.1) was used in organizing the behavioral

data and matching the timecodes.

3.6.2 Raw EEG data

With seven subjects performing each task twice, the total number of EEG data sets was 42. The raw EEG data was analyzed offline, by first filtering through 0–100 Hz bandstop filter, then segmented into 0–150 ms segments from the beginning of each auditory stimuli. These segments were then subjected to artefact rejection, removing any segments with amplitude value exceeding ± 55 μV. Baseline correction was implemented based on the 0–30 ms section of the segments (the silent period of the stimulus, see Figure 3.1). After the baseline correction the segments were averaged to get a single subject, single task average response. Mean of 3191($\pm$S.E.M. 32) segments were averaged per subject. All of these analysis were done with the BrainVision Analyzer.

Two different reference models were calculated in addition to the original FCz-channel reference: grand average reference, where the grand average of all the EEG channels was subtracted from individual channel values; and a horizontal reference, where the sum of TP9 and TP10 channels was subtracted from the individual channel values. Signal-to-noise (SNR) ratios were calculated in Matlab comparing the prestimulus period of 0–30 ms to the 30–150 ms period from the three different reference models to find out which produced the best SNR values.

FFR data

The frequency following response was identified by visual inspection from the averaged single subject responses from the TP9 channel. Across subjects, 6 distinct, consecutive FFR peaks were chosen and the time code for the start and end for this period marked. Rectified mean amplitudes (RMA) were calculated over the FFR region of the responses in Matlab. The mean amplitudes for the first peak of the FFR were also calculated across conditions.

Chapter 4

Results

4.1 Behavioral data

The average hit rates (hits / (hits + misses + false alarms)) across the sets and the conditions are shown in Table 4.1. The hit rates were subjected to 2-way repeated measures ANOVA with set and condition as the within-subject factors. There was no significant difference between sets ($F(1,6) = 0.0012$, $p = 0.9730$) or interaction between sets and conditions ($F(1,6) = 1.0636$, $p = 0.3422$), but there was a significant difference between conditions ($F(1,6) = 11.64$, $p = 0.0143$). Subjects identified targets significantly better in the vowels condition. Because there was no significant difference between the two sets, their combined results are shown in Figure 4.1.

Table 4.1: Mean hit rates across sets and conditions (mean±S.E.M. %)

Set	Expanding rings	Vowels
Set1	84.25±2.17 %	92.31±1.78 %
Set2	81.75±3.78 %	94.76±1.99 %
Sets combined	83.00±2.12 %	93.52±1.33 %

The mean reaction times across sets and conditions are shown in Table 4.2. They were subjected to similar 2-way repeated measures ANOVA as the hit rates, with sets and conditions as the within-subject factors. Results were similar to that of the the hit rates: there

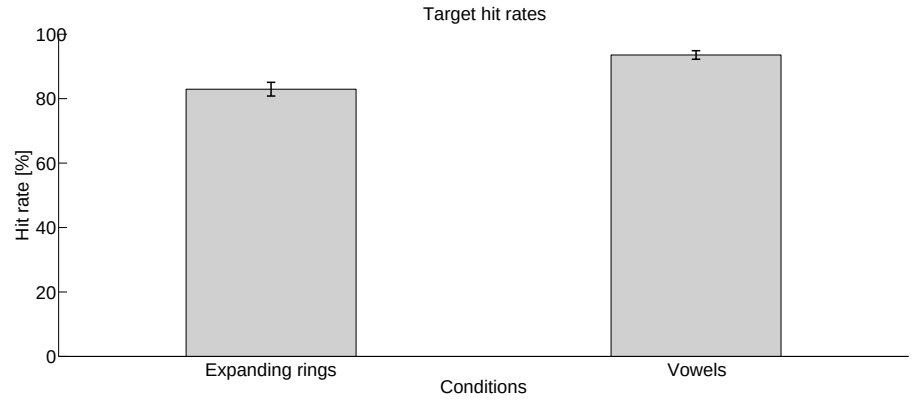

Figure 4.1: The mean hit rate across conditions

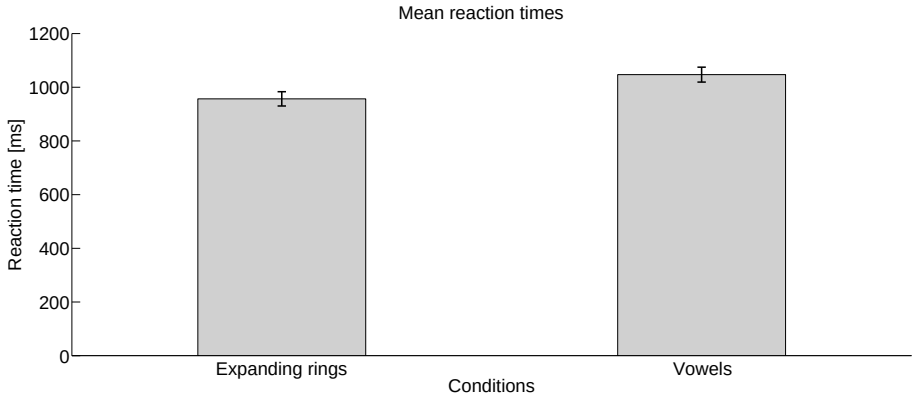

Figure 4.2: The mean reaction times to target stimuli

was no significant difference between sets ($F(1,6) = 0.2079$, $p = 0.6645$) or interaction between sets and conditions ($F(1,6) = 2.0387$, $p = 0.2033$), but there was a significant difference between conditions ($F(1,6) = 39.71$, $p < 0.001$). Subjects reacted significantly slower to targets when identifying vowels (lipreading). Because there was no significant difference between the two sets, their combined results are shown in Figure 4.2.

Table 4.2: Mean reaction times across sets and conditions (mean±S.E.M. ms)

Set	Expanding rings	Vowels
Set1	951.14±39.02 ms	1060.29±43.95 ms
Set2	962.43±39.14 ms	1033.57±36.56 ms
Sets combined	956.79±26.6 ms	1046.93±27.1 ms

The mean discrimination indexes (d-prime) across sets and condition are shown in Table 4.3. They were subjected to 2-way repeated measures ANOVA with set and condition

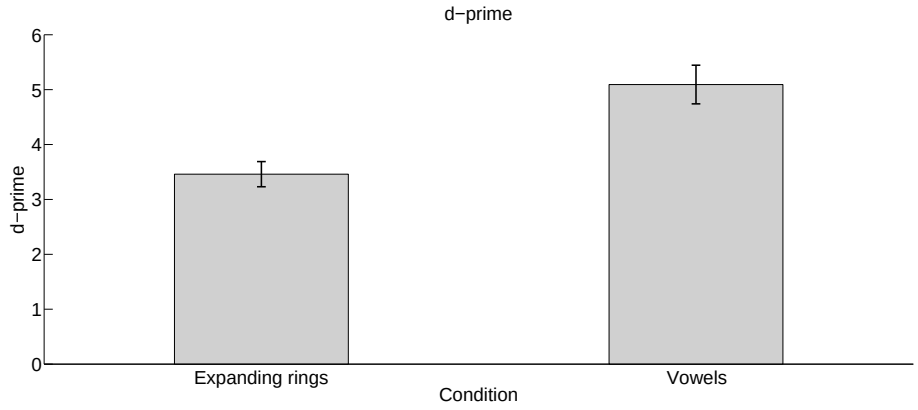

Figure 4.3: The mean discrimination index (d-prime) across conditions

as the within-subject factors. ANOVA showed signifcant difference between conditions ($F(1,6) = 43.10$, $p < 0.001$), but no significant difference between sets ($F(1,6) = 0.99$, $p = 0.3572$) or interactions between set and condition ($F(1,6) = 4.54$, $p = 0.1006$). This was in accordance with the hit rate results. Because there was no significant difference between the two sets, their combined results are shown in Figure 4.3.

Table 4.3: Mean discrimination indexes (mean±S.E.M.)

Set	Expanding rings	Vowels
Set1	3.66±0.40	4.48±0.55
Set2	3.26±0.23	5.70±0.34
Sets combined	3.46±0.23	5.09±0.35

4.2 EEG results

When analyzing all the 42 FCz-referenced EEG datasets 45 % of the maximum SNRs were achieved on the TP9 channel (TP10: 24 %) (see Figure 4.4). In the average-referenced EEG datasets 40 % of the maximum SNRs were on the TP9 channel (TP10: 19 %) (see Figure 4.5). In the horizontal-referenced dataset 21 % of the maximum SNRs were on channel Cz (Fp2, Fz, FC1 and CP2: 7 %) (see Figure 4.6). The SNR results suggested the original FCz-referenced montage to be best suited for further analysis, as it gave the best SNR values for the appropriate channel (TP9).

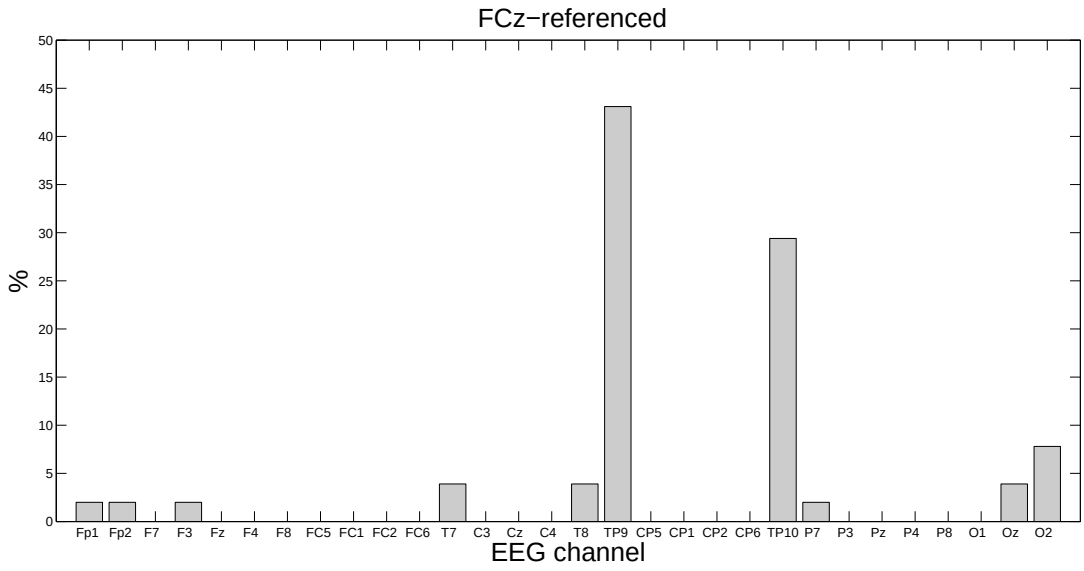

Figure 4.4: FCz-reference: best SNR division between channels in percentages

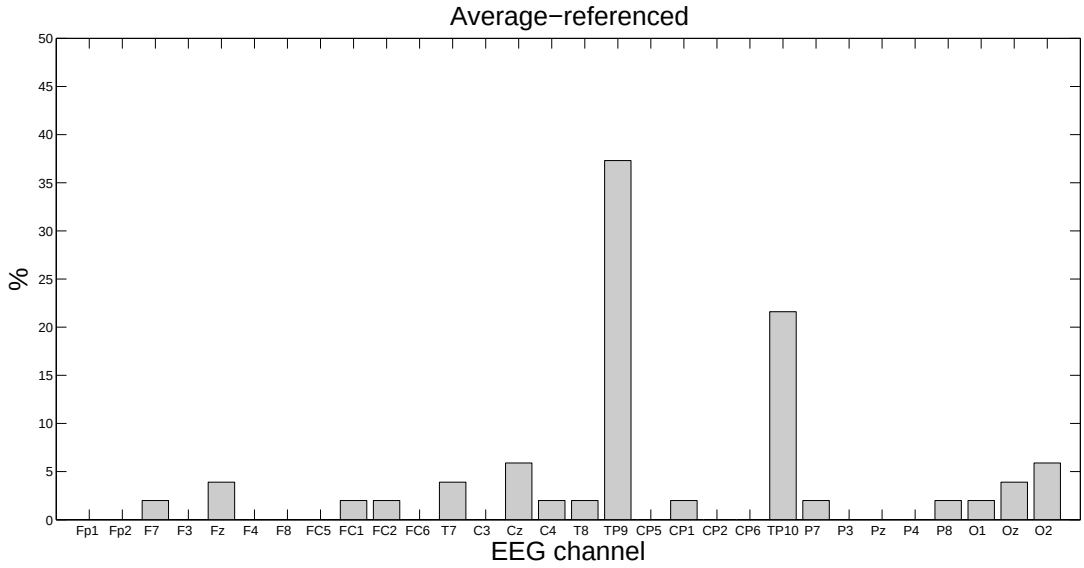

Figure 4.5: Average-reference: best SNR division between channels in percentages

4.2.1 FFR

An example of the visual identification of the FFR portion is shown in Figure 4.7. FFR durations were calculated from peak-to-peak. Durations for set1 over the conditions can be seen in Table 4.4 and for set2 in Table 4.5. The durations of the identified FFR portions were subjected to 2-way repeated measures ANOVA with set and condition as within-subject factors. The results indicated no significant differences between any factor com-

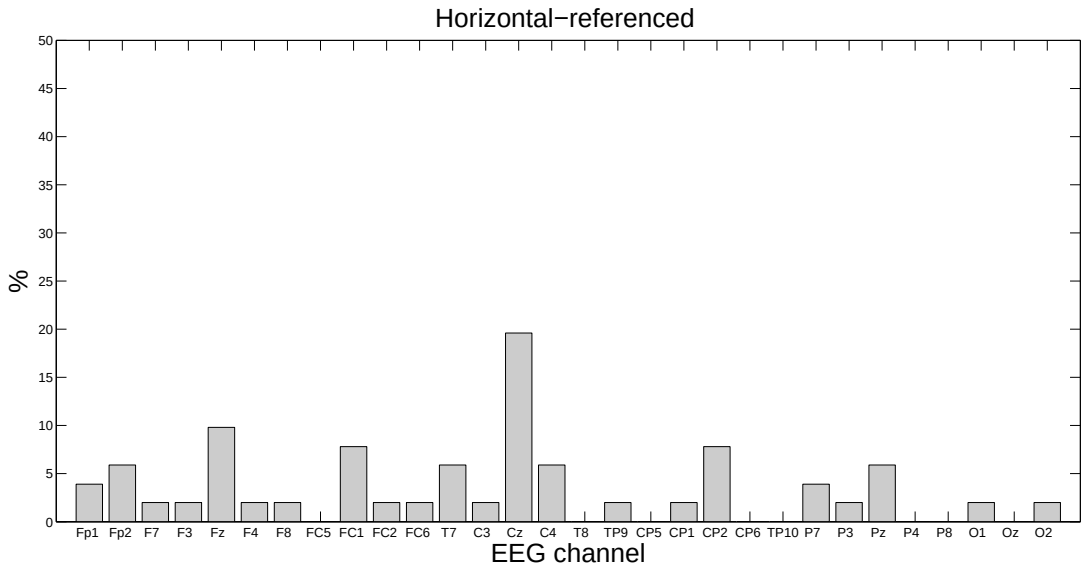

Figure 4.6: Horizontal-reference: best SNR division between channels in percentages

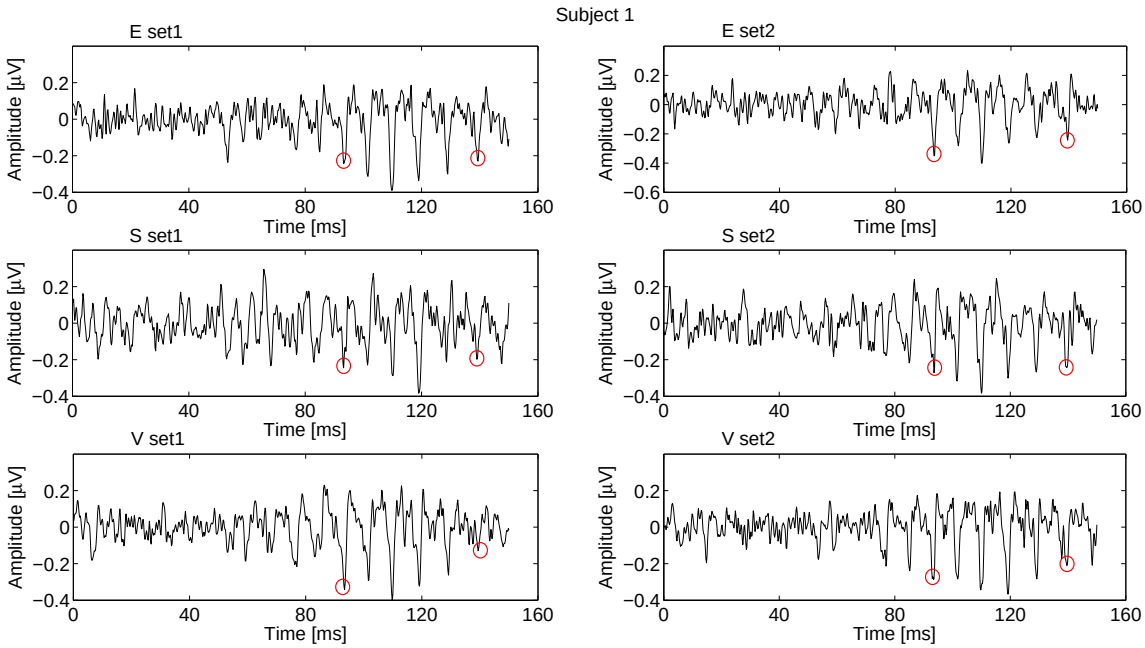

Figure 4.7: FFR identification from the auditory brainstem response of a single subject

binations (sets, F(1,6) = 1.28, p = 0.3014; conditions F(2,6) = 1.59, p = 0.2443 and sets and conditions (F(1,2) = 0.44, p = 0.6550). Because no significant differences between sets were found, the mean durations were counted over both sets, see Table 4.6 and Figure 4.8.

Table 4.4: Condition specific FFR durations, set1 (ms)

Subject	Expanding rings	Still face	Vowels
1 (set1)	46.2	45.8	46.0
2 (set1)	43.0	45.6	43.2
3 (set1)	46.0	46.8	33.2
4 (set1)	45.4	45.8	45.8
5 (set1)	46.0	45.8	46.2
6 (set1)	46.0	45.6	46.0
7 (set1)	46.2	46.2	46.0
Mean	45.543	45.943	43.771
S.E.M	0.403	0.150	1.672

Table 4.5: Condition specific FFR durations, set2 (ms)

Subject	Expanding rings	Still face	Vowels
1 (set2)	46.2	46.0	46.2
2 (set2)	42.6	46.4	43.6
3 (set2)	46.8	46.4	43.0
4 (set2)	46.2	46.8	45.8
5 (set2)	45.8	46.2	46.0
6 (set2)	46.0	46.4	44.2
7 (set2)	46.0	45.8	46.0
Mean	45.657	46.286	44.971
S.E.M	0.484	0.113	0.467

Table 4.6: Condition specific FFR durations, both sets (ms)

	Expanding rings	Still face	Vowels
Mean	45.6	46.11	44.37
S.E.M	0.327	0.108	0.916

Rectified mean amplitudes (RMA) were calculated over the identified FFR portion for each subject, set and condition. Table 4.7 shows the RMA values for set1 over all subjects and conditions and Table 4.8 shows the corresponding values for set2. The RMA values

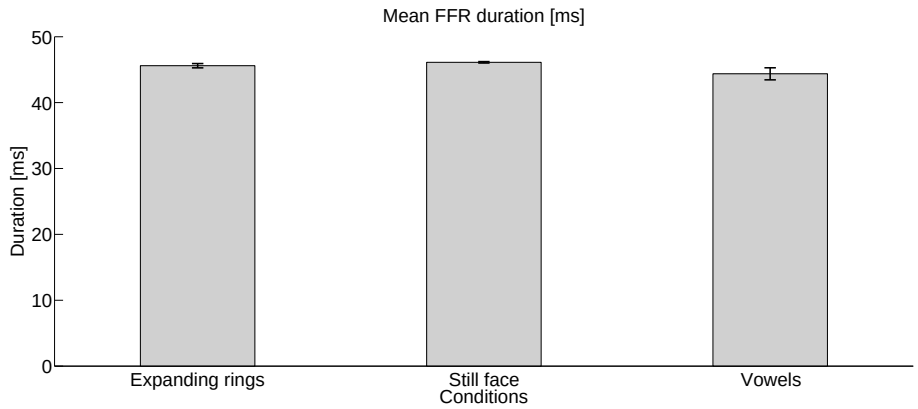

Figure 4.8: Mean durations of the FFR portion of the responses across conditions

were subjected to 2-way repeated measures ANOVA with set and condition as the within-subject factors, and no significant differences were found between any combination of factors (sets, $F(1,6) = 1.77$, $p = 0.2311$; conditions $F(2,6) = 1.15$, $p = 0.3480$ and sets and conditions ($F(1,2) = 0.50$, $p = 0.6164$).

Because there was no significant difference between the two sets, the mean RMA values were calculated over both sets (see Table 4.9). The results are shown in Figure 4.9.

Table 4.7: Condition specific RMA values of the FFR, set1 (μV)

Subject	Expanding rings	Still face	Vowels
1 (set1)	0.104	0.096	0.099
2 (set1)	0.103	0.097	0.102
3 (set1)	0.351	0.151	0.143
4 (set1)	0.344	0.190	0.130
5 (set1)	0.112	0.261	0.115
6 (set1)	0.105	0.176	0.108
7 (set1)	0.097	0.081	0.100
Mean	0.148	0.131	0.103
S.E.M.	0.035	0.024	0.008

The mean peak amplitude values of the first identified peak of the FFR were analyzed with 2-way repeated measures ANOVA with set and condition as within-subject factors. No significant differences were found between sets ($F(1,6) = 0.04$, $p = 0.8494$) or conditions ($F(2,6) = 2.51$, $p = 0.1227$), but there was significant interaction between sets and conditions ($F(1,2) = 5.53$, $p = 0.0198$). The peak amplitude values for set1 are shown in

Table 4.8: Condition specific RMA values of the FFR, set2 (μV)

Subject	Expanding rings	Still face	Vowels
1 (set2)	0.092	0.088	0.085
2 (set2)	0.137	0.107	0.101
3 (set2)	0.123	0.098	0.101
4 (set2)	0.145	0.145	0.077
5 (set2)	0.079	0.186	0.093
6 (set2)	0.093	0.074	0.087
7 (set2)	0.084	0.089	0.071
Mean	0.133	0.132	0.099
S.E.M.	0.035	0.018	0.007

Table 4.9: Condition specific mean RMA values of the FFR, combined sets (μV)

	Expanding rings	Still face	Vowels
Mean	0.141	0.131	0.101
S.E.M.	0.024	0.015	0.005

Table 4.10 and for set2 in Table 4.11. Both sets are shown in Figure 4.10. Additional paired Student's *t*-tests results showed significant difference between [expanding rings set1], [vowels set2] (p = 0.0490) and close to significant differences between [still face set1], [vowels set 2] (p = 0.0893) and [still face set2], [vowels set2] (p = 0.0880).

Cross-correlations between the ABRs and the auditory stimulus were calculated in Matlab to find the latency of the response. The stimulus was shifted in time to find the maximum

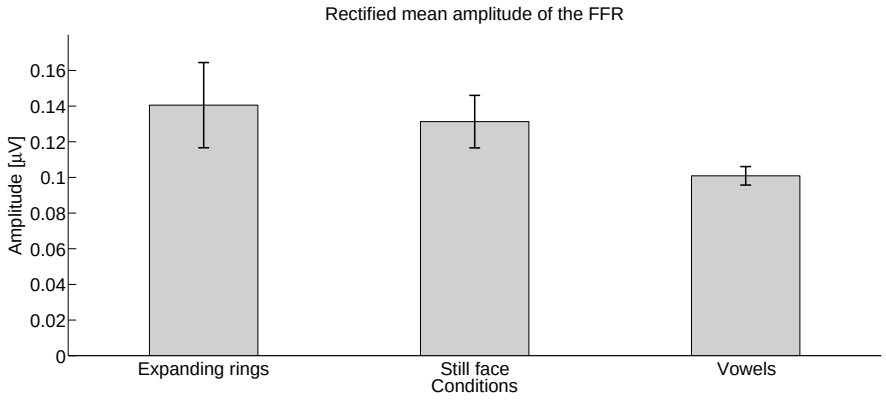

Figure 4.9: Mean RMA values of the FFR across conditions

Table 4.10: Condition specific mean amplitude values of the first FFR peak, set1 (μV)

Subject	Expanding rings	Still face	Vowels
1 (set1)	-0.244	-0.245	-0.341
2 (set1)	-0.428	-0.507	-0.135
3 (set1)	-0.283	-0.563	-0.287
4 (set1)	-0.298	-0.229	-0.330
5 (set1)	-0.317	-0.305	-0.366
6 (set1)	-0.325	-0.374	-0.293
7 (set1)	-0.204	-0.255	-0.162
Mean	-0.300	-0.354	-0.273
S.E.M	0.025	0.047	0.031

Table 4.11: Condition specific mean amplitude values of the first FFR peak, set2 (μV)

Subject	Expanding rings	Still face	Vowels
1 (set2)	-0.351	-0.272	-0.286
2 (set2)	-0.735	-0.964	-0.282
3 (set2)	-0.230	-0.501	-0.140
4 (set2)	-0.238	-0.276	-0.335
5 (set2)	-0.274	-0.244	-0.181
6 (set2)	-0.217	-0.437	-0.185
7 (set2)	-0.184	-0.242	-0.134
Mean	-0.318	-0.419	-0.220
S.E.M	0.067	0.091	0.028

correlation between the response and stimulus, indicating the timing of the response. Here the effects of the distance the sound had to travel first were taken into account (see Chapter 5), because the segmentation of the raw EEG data was based on the stimulus triggers, but the actual sound took approximately 3.6 ms from the trigger to travel to the subject. This propagation time has been subtracted from all of the reported latencies.

The latencies were subjected to 2-way repeated measures ANOVA with set and condition as within-subject factors. The results indicated no significant differences between any factor combinations (sets, $F(1,6) = 0.13$, $p = 0.7345$; conditions $F(2,6) = 2.33$, $p = 0.1393$ and sets and conditions ($F(1,2) = 1.55$, $p = 0.2516$). Because no significant differences between sets were found, the mean latencies were counted over both sets (see Figure 4.11 and Table 4.12).

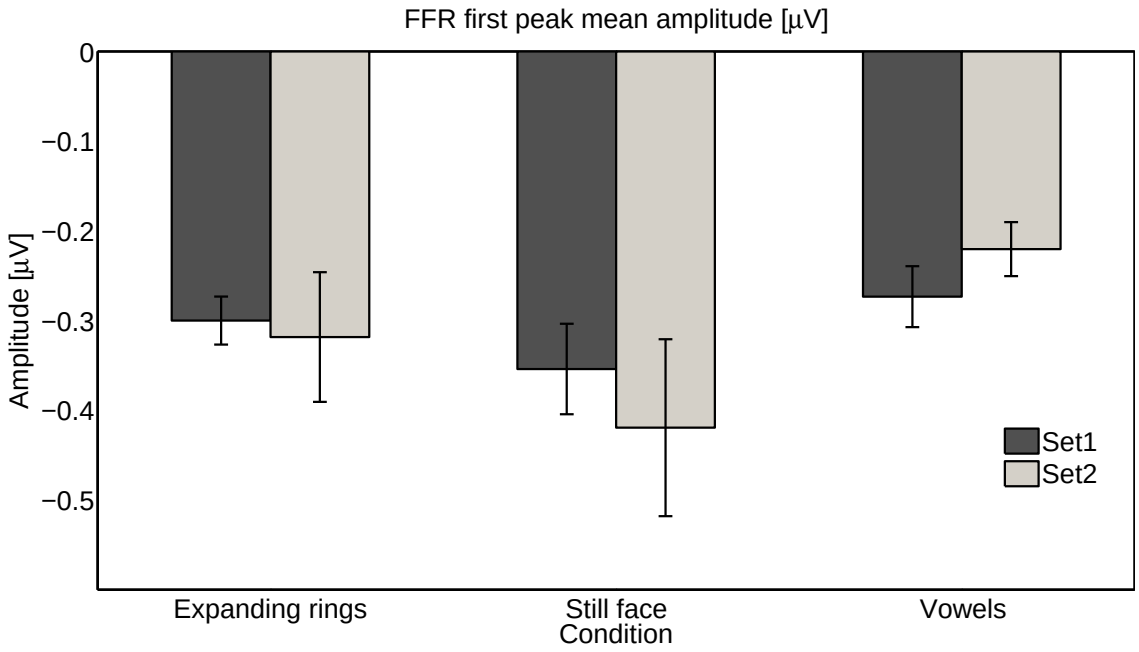

Figure 4.10: Mean amplitude values of the first peak of the FFR across conditions and sets

Table 4.12: Mean response latencies (mean±S.E.M. ms)

	Expanding rings	Still face	Vowels
Mean	19.93	22.93	18.6
S.E.M	1.30	0.25	2.04

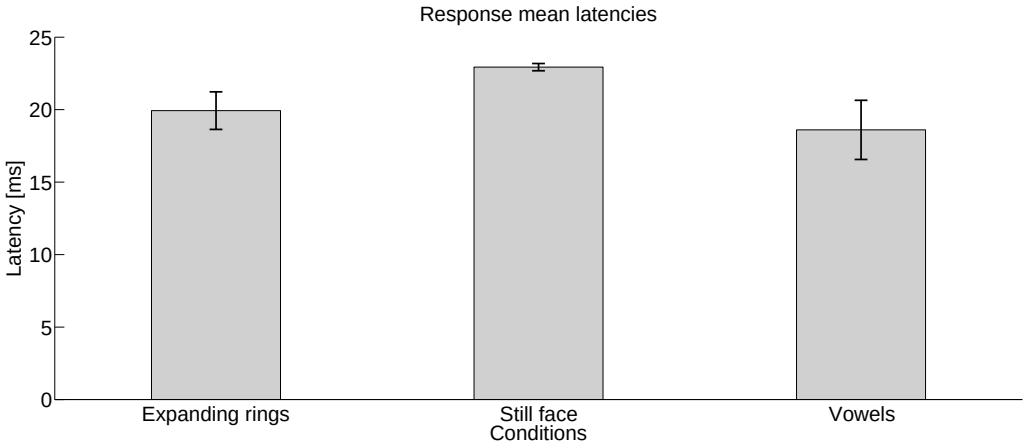

Figure 4.11: Mean latencies of the ABR response across conditions

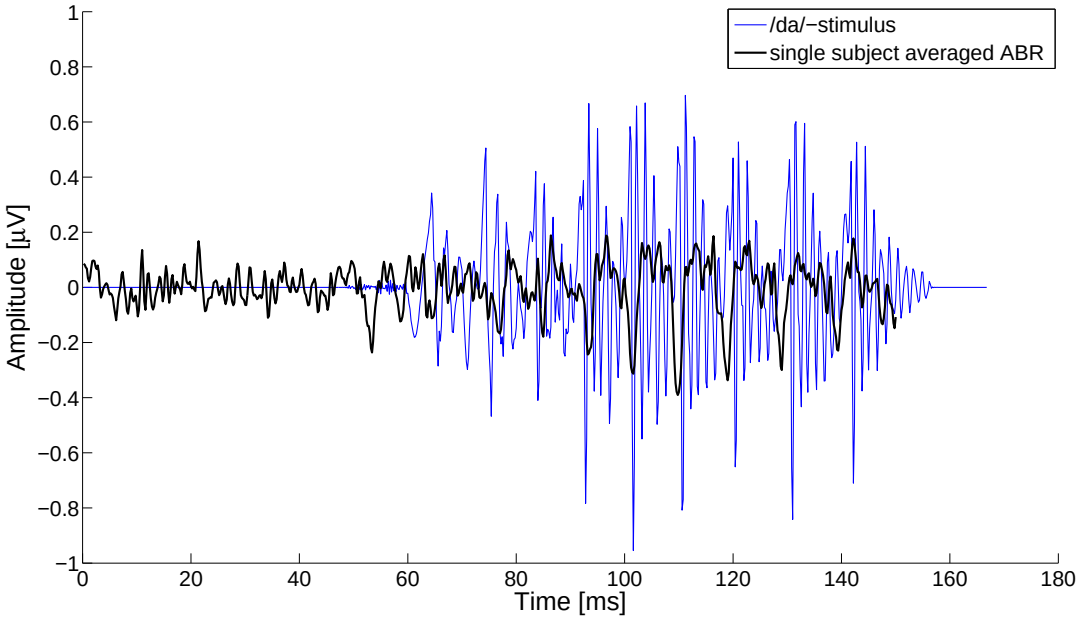

Figure 4.12: The grand average single subject ABR (E-condition) overlaid on the stimulus, with the stimulus shifted in time to correspond to the latency of the response

Chapter 5

Discussion

5.1 Summary of results

The behavioral results (hit rate, d-prime) show that the subjects succeeded better in identifying the vowel targets than the expanding rings. Hit rates differed slightly from what Kauramäki et al. (2010) reported, as our subjects had a significantly better hit rate in the vowels / lipreading task. In Pekkola et al. (2005) the subjects performed marginally better in the vowel condition, so the differences to Kauramäki et al. (2010) might be due to the subject group, differences in attention or experiment conditions.

The reaction times were longer in the vowels condition than in the expanding rings condition. This result was in accordance with Kauramäki et al. (2010).

When analyzing the frequency following response portion of the auditory brainstem response, no statistically significant effects were found in the duration or the rectified mean amplitude of the FFR. When analyzing the mean amplitudes of the first peak of the FFR, no statistically significant overall effect was observed over the conditions. However, there was a statistically significant effect within the two separate task sets between the first expanding rings set and the second vowels set. The mean peak amplitude was significantly larger in the expanding rings condition within these two sets. Close to significant results

were observed between the both still face -sets in respect to the second vowels set, with larger mean peak amplitudes in the still face sets.

No significant effect was observed in the latencies of the auditory brainstem response across conditions.

5.1.1 Effects of lipreading on the FFR

We observed no effects between conditions on the length of the FFR. There were no statistically significant differences between conditions in the RMA of the FFR. However, as can be seen from Figure 4.9, the vowels condition has the smallest RMA value. Musacchia et al. (2006) observed significant amplitude suppression effect on the early portion of the ABR in audiovisual speech condition vs. auditory-only condition. As the auditory stimuli presented in our experiment were not in synchrony with the visual stimuli, it is possible that if lipreading causes short-term suppression in the FFR, the effect gets lost in the averaging process.

The only statistically significant result was in the FFR first peak amplitude within the two separate task sets. The mean peak amplitude was significantly larger in the expanding rings set1 than in the vowel/lipreading set2. Close to significant results were observed between the both still face sets in respect to the second vowels set, with larger mean peak amplitudes in the still face sets. Visual attention has been proved to increase response amplitudes (Valtonen et al., 2003), which would suggest the first peak results stem from some other source than attention, because the still face condition, where attention was not controlled with a task produced the largest amplitudes. The results obtained here are inconclusive, but the trend seen in the amplitudes could suggest tentative audiovisual integration effects in the FFR peak amplitudes.

5.2 Experiment improvement suggestions

When looking at the experiment described in this thesis purely from a technical point of view, we did not succeed in reliably recording the first I-V waves of the auditory brain-stem response. However, a robust representation of the frequency following response was achieved.

The biggest issue with the equipment was the relatively low sampling rate of the EEG-equipment, making it ill-suited for measuring the ABR. If focusing on the first I-V waves of the ABR (see Figure 2.9), where the changes are rapid within the first 5 – 10 ms, 5000 Hz sampling rate is not enough to reliably depict the response. Hyde (1985) suggest a sampling rate of at least 10 kHz for ABR recording. For example Musacchia et al. (2006) used 20 kHz sampling frequency in their ABR study.

The second largest issue was using a standalone speaker instead of ear-insert phones. Using a speaker introduced some additional latency in the propagation of the sound wave to the subject which could not be monitored or kept constant between subjects. Had we been able to reliably record and identify the I-V waves of the ABR, the propagation time of sound over air (see Table 5.1) would have added an uncertainty factor to the latencies where fractions of milliseconds are significant: difference of mere 5 cm in the distance would introduce 0.14 ms more latency in the propagation. Additionally there were some doubts on whether the sound signal still coupled electrically to EEG channels, despite the careful measures taken to minimize this artefact.

Table 5.1: Propagation time of a sound wave in 25°C (346.1 ms/s)

Distance (m)	1.15	1.2	1.25	1.3
Propagation time (ms)	3.32	3.47	3.61	3.76

Having only one size of the ACTIcap also presented some problems, as the size of the head was not a criterion of subject selection. Thus there were some issues with the cap being too large or too small, ultimately leading to the remocal of one subject from the analysis (see Section 3.2). If the head size of the subject was known beforehand, a proper

sized cap could be prepared in advance.

A better attention control could be introduced to the still face condition, for example by introducing a small portion of deviant auditory stimuli that subjects should count (Musacchia et al., 2006). This would make the conditions equal on attentional basis.

Appendix A

Individual ABR data on channel TP9

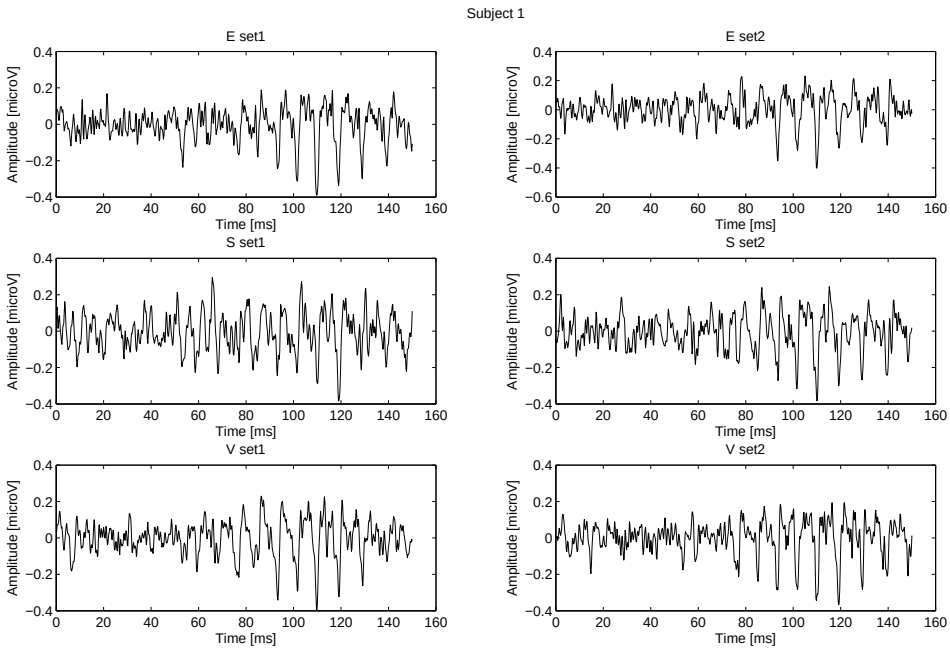

Figure A.1: Grand averaged auditory brainstem responses across conditions, subject 1

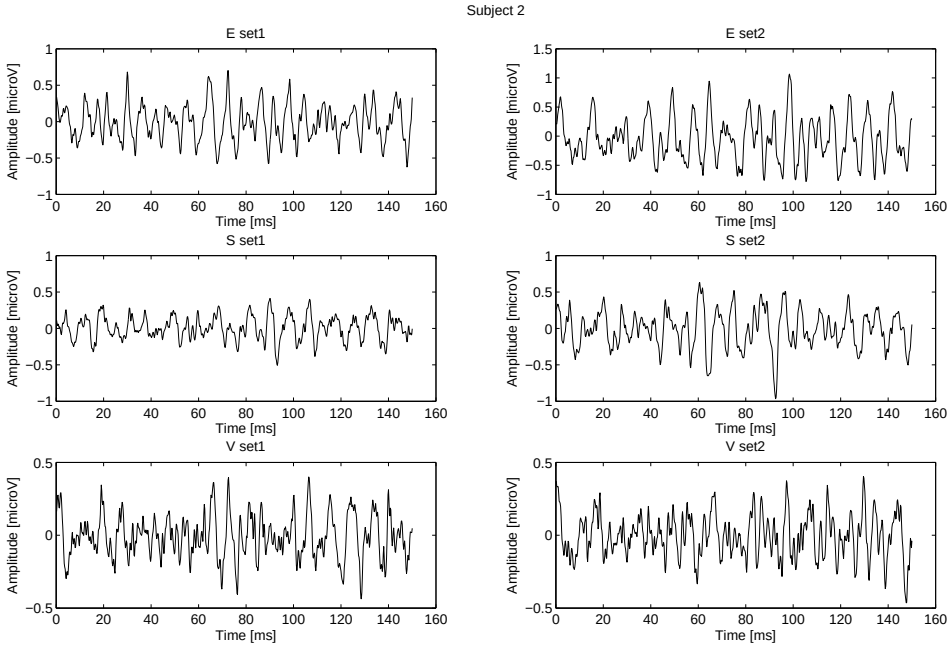

Figure A.2: Grand averaged auditory brainstem responses across conditions, subject 2

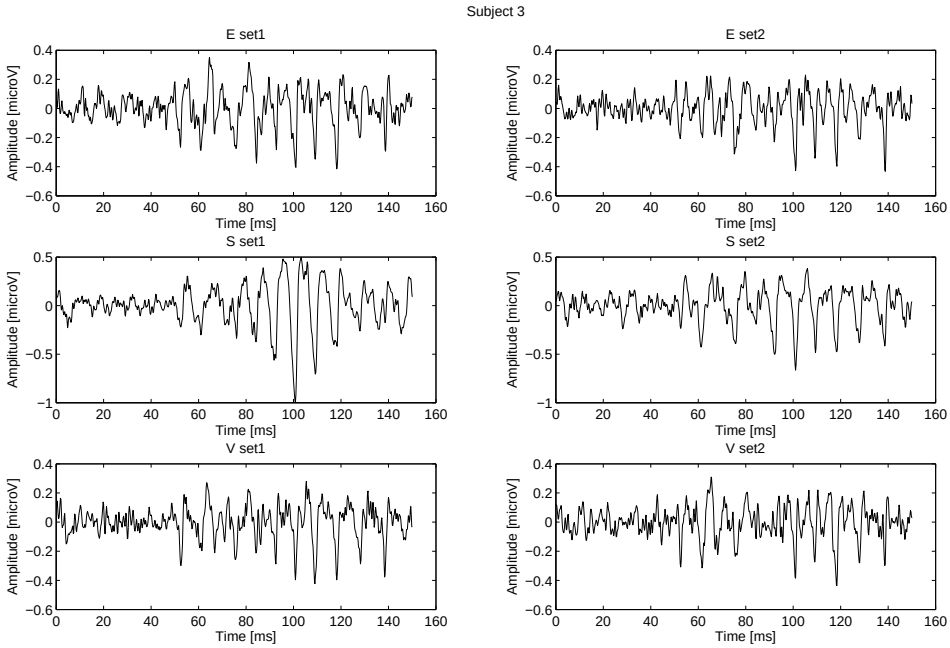

Figure A.3: Grand averaged auditory brainstem responses across conditions, subject 3

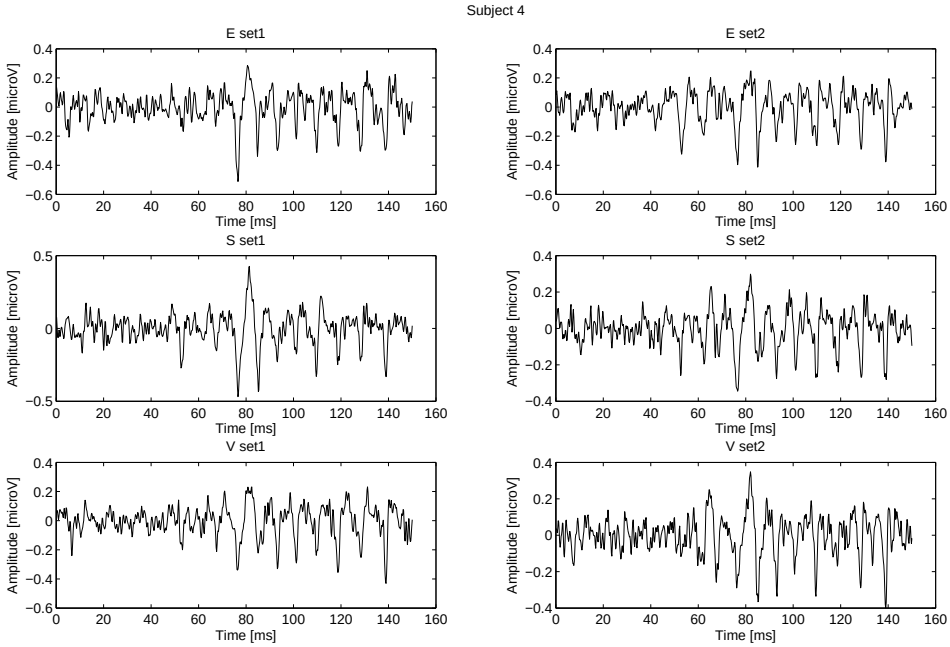

Figure A.4: Grand averaged auditory brainstem responses across conditions, subject 4

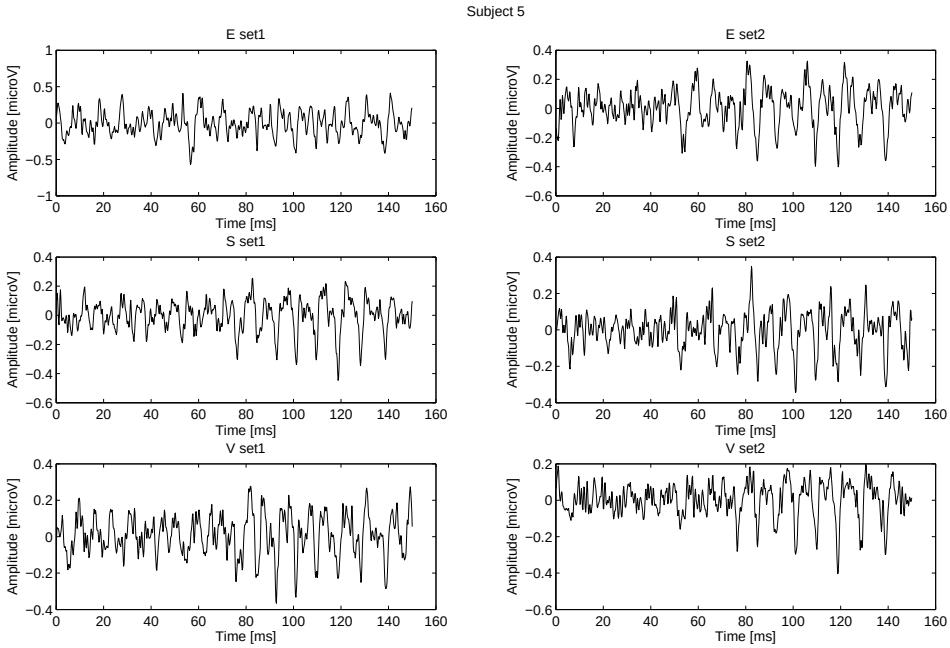

Figure A.5: Grand averaged auditory brainstem responses across conditions, subject 5

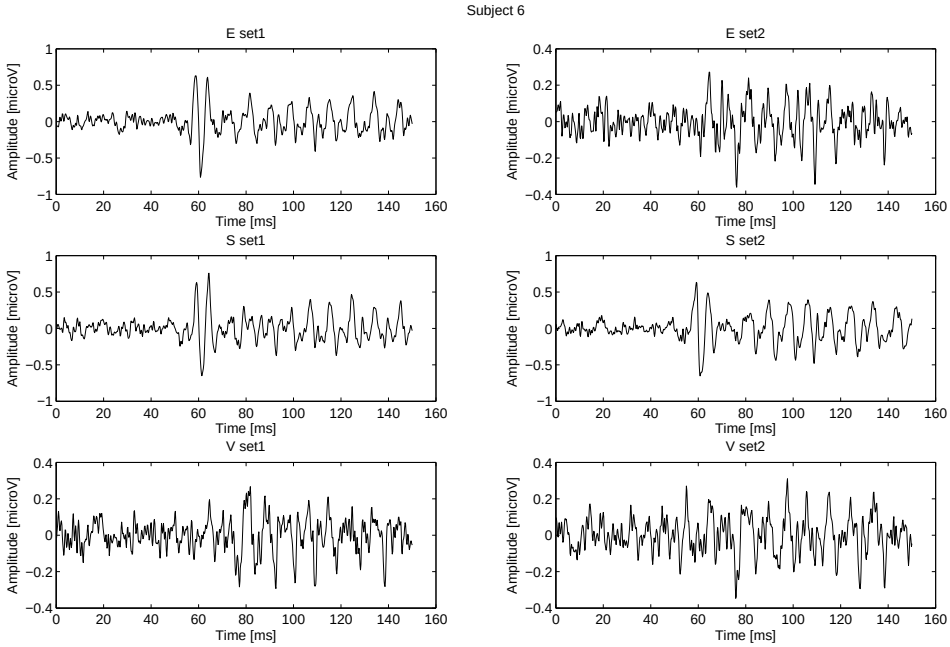

Figure A.6: Grand averaged auditory brainstem responses across conditions, subject 6

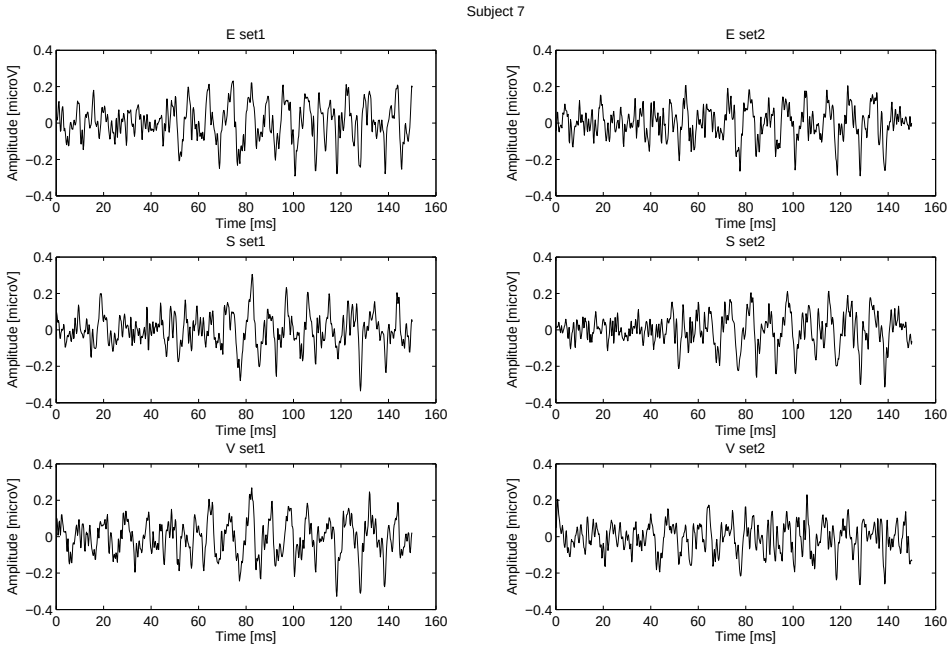

Figure A.7: Grand averaged auditory brainstem responses across conditions, subject 7

Bibliography

Bear, M. F., Connors, B. W., and Paradiso, M. A. (2001). *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, 2nd edition.

Besle, J., Bertrand, O., and Giard, M.-H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*, 258:143 – 151.

Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8):2225 – 2234.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276:593 – 596.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., and Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clinical Neurophysiology*, 112:758 – 767.

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind*. W. W. Norton & Company, Inc., 2nd edition.

Goldstein, E. B. (2002). *Sensation and Perception*. Wadsworth, 6th edition.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164 – 171.

Hyde, M. L. (1985). *The Auditory Brainstem Response*, chapter Instrumentation and Signal Processing, pages 33 – 48. Taylor & Francis Ltd, original publisher College-Hill Press.

Jacobson, J. T., editor (1985). *The Auditory Brainstem Response*. Taylor & Francis Ltd, original publisher College-Hill Press.

Johnson, K. L., Nicol, T. G., Zecker, S. G., and Kraus, N. (2007). Auditory brainstem correlates of perceptual timing deficits. *Journal of Cognitive Neuroscience*, 19(3):376 – 385.

Karjalainen, M. (1999). *Kommunikaatioakustiikka*. Report 51, Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing.

Kauhanen, L., Jylänki, P., Lehtonen, J., Rantanen, P., Alaranta, H., and Sams, M. (2007). Eeg-based brain-computer interface for tetraplegics. *Computational Intelligence and Neuroscience*, 2007:1 – 11.

Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., and Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *The Journal of Neuroscience*, 30(4):1314 – 1321.

Klucharev, V., Möttönen, R., and Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, 18(1):65 – 75.

Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism – Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746 – 748.

Mildner, V. (2008). *The Cognitive Neuroscience of Human Communication*. Taylor Francis Group, LLC.

Musacchia, G., Sams, M., Nicol, T., and Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, 168(1 – 2):1 – 10.

Niedermeyer, E. and da Silva, F. L., editors (1999). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 4th edition.

Nunez, P. L. (2006). *Electric Fields of the Brain: the Neurophysics of EEG*. Oxford University Press, 2nd edition.

Pekkola, J., Ojanen, V., Auntti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., and Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*, 16(2):125 – 128.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., Lamantia, A.-S., Mcnamara, J. O., and Williams, S. M. (2004). *Neuroscience*. Sinauer Associates, Inc, 3rd edition.

Valtonen, J., May, P., Mäkinen, V., and Tiitinen, H. (2003). Visual short-term memory load affects sensory processing of irrelevant sounds in human auditory cortex. *Cognitive Brain Research*, 17(2):358 – 367.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4):1181 – 1186.

Yost, W. A. (2000). *Fundamentals of Hearing: An Introduction*. Academic Press, Elsevier Science, 4th edition.